Inverse Scattering and Shape Reconstruction

A thesis presented for the degree of
Doctor of Philosophy
in Electrical & Electronic Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

by
Ross David Murch
B. E. (Hons 1)
April 1990

Abstract

Investigations of new and improved solutions to inverse problems are considered. Three of the solutions are concerned with inverse scattering. The other two solutions deal with reconstructing binary images from few projections and determining the shape and orientation of a three-dimensional object from silhouettes. In addition, a review of solutions to direct and inverse scattering problems is presented.

An inverse scattering algorithm for reconstructing variable refractive index distributions is examined. The inversion algorithm is based on an expression for the wave function which explicitly incorporates the inverse scattering data. It is claimed that this considerably increases the efficiency of the algorithm. The algorithm is implemented in two-dimensional space and examples of reconstructions of objects from computer-generated scattering data are presented.

The problem of determining the shape of a two-dimensional impenetrable obstacle from a set of measurements of its far-field scattering amplitude is considered. The problem is formulated as a non-linear operator equation which is solved by an iterative method. The use of the null-field method to solve the direct problem leads to efficient evaluation of the Fréchet derivative of the non-linear operator. Computational implementations confirm the numerical accuracy of the algorithm.

An extension to the Rayleigh-Gans (Born) approximation is examined. The extension involves incorporating a high frequency approximation to the wave field into the conventional Rayleigh-Gans (Born) approximation. Numerical implementation of an algorithm based on this extension to the Rayleigh-Gans (Born) approximation indicates that its reconstruction accuracy is generally superior to that of the conventional Rayleigh-Gans (Born) approximation.

Efficient algorithms for reconstructing a binary cross-section (each of whose pixel amplitudes is either zero or unity) from few one-dimensional projections are introduced and illustrated by example. It is shown that only two projections are needed to reconstruct a convex cross-section. Non-convex cross-sections need more projections but far fewer than are necessary to reconstruct grey-scale images. When presented with noisy one-dimensional projections, the algorithms remain useful, although their performance improves with the number of given projections.

Determination of a three-dimensional object's shape and orientation from its silhouettes is studied, on the understanding that the relative orientations of the given silhouettes are unknown *a priori*. The result of this study is an algorithm which could be suitable for incorporation into a robot's vision system. The algorithm is based on a method for determining the orientation of an object from its two-dimensional projections. To overcome the reduced information content of silhouettes as compared with two-dimensional projections, a self consistency check is introduced. Numerical implementations of the algorithm confirm that it can generate usefully accurate estimates of the orientations and shapes of technologically non-trivial objects.

Acknowledgements

I wish to sincerely thank my supervisor, Professor R. H. T. Bates, for his enthusiasm, inspiration and persistence. I have benefited greatly both academically and personally from his supervision, and am truly indebted to him.

Collaborators Dr T. John Connolly, Dr Susanne M. Dale, S. Tony Enright, Bruce P. McGregor, Vaughan A. Smith, Dr David G. H. Tan and Dr David J. N. Wall have also been invaluable. Their expertise and imagination have generated excitement in my research. The past and present members of Room 7, my study room, are also thanked for their friendliness and honesty. I also express gratitude to the people studying and working around me as they have provided support and creative discussions. The generous financial help given by Telecom corporation of New Zealand has also been much appreciated.

I especially thank my family and friends for always being there to lean on. Of particular mention is Dr Alan Murch, my brother, whose rare perspective on life has always enlightened me. Andrea McBride and Bruce McGregor, two very special and dependable friends, have also made my studies very memorable. Their distracting influence has kept me sane, I think!

Contents

Preface

Humans have an insatiable desire to discover more and more about the physical world. It is only during the past three to four millennia, however, that the knowledge gained about the physical world has been systematically recorded. The consequent changes this knowledge has brought to our society have been tremendous. Our acquired knowledge, though, is only useful because it rests on objective observation of the physical world.

The more direct our interaction is with the environment, the better pleased we humans are with our observations. Our natural propensity is to touch, and peer into and around, everything we can get close to. The senses of hearing and sight allow us to guess at the nature of objects beyond our physical reach. We are able to see stars in the night sky and hear encroaching thunder storms which are impossible to get close to. We desire to learn even more about such objects than is feasible by our unaided senses. Telescopes, magnifiers and amplifiers allow our senses to have much greater sensitivity and resolution than is otherwise possible. Even in these cases, our senses are vital for our observations. It is consequently reasonable to classify information obtained through our unaided and aided senses as having been acquired directly.

There are also plenty of objects which our senses, no matter how sensitive they are nor how enhanced they may be, cannot gather sufficient information about to satisfy our needs. Examples of some objects are precious minerals in the ground, internal body organs and fish in the deep sea. Our aided or unaided senses then fail because there is no light, sound, smell, taste or contact to perceive the objects by. So, our only recourse is to use instruments which can perceive disturbances or effects outside the range of our senses. We classify information gathered about objects in this way as having been acquired indirectly.

The methods by which we gather information indirectly are many and varied. Some terms for them are remote probing, remote sensing and imaging, all of which involve unravelling causes hidden within the acquired information (Sabatier, 1978; Baltes, 1980). Any requirement to uncover such a cause from information of this kind is often called an inverse problem. It has become conventional to partition information gathering requirements into direct and inverse problems (Sabatier, 1983; Santosa *et al.*, 1984). It is important to understand the difference between them. For instance, consider the sound made by a musical instrument. The direct problem would be to calculate the sound engendered by the instrument given that you knew all about the instrument and how it was played. In contrast, the inverse problem would be to determine the details of the instrument from only the sound emanating from it. The inverse problem arises in a large number of disciplines. Some of these are medical imaging, geophysics, radar, sonar, non-destructive testing and electromagnetic imaging. Many of these inverse problems frequently involve wave-like emanations. It has become common to classify

inverse problems involving wave-like emanations as either inverse scattering or inverse source problems (Devaney and Sherman, 1982). Inverse scattering problems are those in which probing wave-like emanations are used to gather information about objects. Such problems occur in for example computed tomography (Herman, 1983), ultrasonic imaging (Wells, 1977) and geophysics (Schoenberger, 1984). The inverse source problem occurs when emanations from the object of interest are used to gain information about it. It arises in many fields, as diverse as nuclear medicine (Nudelman and Patton, 1980) and astronomy (Craig and Brown, 1986).

The concern of this thesis is with the inverse problem. My interest in this problem lies in devising new or improved methods so that more intricate and precise observations of objects can be made indirectly. Several new and improved techniques are introduced here in the areas of computer vision, reconstruction from projections and inverse scattering. Computer vision is becoming increasingly important in the development of robots, which must be able to locate, recognize and orientate objects efficiently in order to perform useful tasks (Billingsley, 1985). To this end, I have developed a technique (introduced in Chapter 7) which allows an object's shape and orientation to be found from a finite number of its (unoriented) silhouettes. There are also situations of practical importance in computer vision applications in which only a very limited number of views of a body can be recorded. I have developed an algorithm (introduced in Chapter 6) which enables cross-sections of objects of constant density to be reconstructed from as few as two projections. I have also been concerned with improving inverse scattering techniques which estimate properties of scatterers from measurements of scattered wave fields. Three new techniques are presented in Chapter 5.

Each of the following paragraphs summarises one of the chapters in the thesis. All items termed "original" relate to aspects of my PhD research. So, results said to be original have been obtained by myself, either alone or together with others. Whether or not any particular result has involved collaboration is made clear by the references quoted in the particular section(s) of the relevant chapter.

Chapter 1 serves to introduce, in descriptive terms, some solutions to inverse problems which are in practical use today. It also highlights some of the limitations of these solutions, so that desirable improvements to them can be identified for research. Terminology to describe inverse problems in a wide range of situations is introduced. The direct and inverse problems are stated formally and general techniques for their solution are discussed. The dimensionality of the space in which the inverse problem is set also greatly effects the difficulty experienced in solving it. It is indicated why that inverse problems that are set in two or more dimensions are both conceptually and technically more intricate and difficult to solve than one-dimensional problems.

Many physical processes manifest themselves as scalar linear wave fields under a wide range of scientifically and technologically interesting situations. This allows many inverse and direct scattering problems to be treated in a unified manner. Chapter 2 introduces time- and frequency-domain equations describing wave fields. It is explained why there is a preference in the literature to pose inverse problems in the frequency domain. Supplementary equations and terms which are useful for descriptions of wave fields are also introduced.

Solutions to direct scattering problems which are useful (or potentially useful) for solving inverse scattering problems are reviewed in Chapter 3. Approximate and exact methods are examined. An original ray-tracing algorithm is presented in this chapter.

Solutions to inverse scattering problems are reviewed in Chapter 4. Mathematical as well as numerical aspects are treated. The review is divided into four subsections covering what I call exact explicit, approximate explicit, exact implicit and approximate implicit solutions to inverse problems. The uniqueness, existence and stability of solutions are discussed in appropriate detail.

Chapter 5 introduces three original techniques for solving inverse scattering problems. The first two of these techniques are based on exact descriptions of wave fields. One of them is a method for reconstructing the spatially varying refractive index distribution inside an object. The other technique enables the shape of an impenetrable object to be reconstructed. The last technique introduced in this chapter is based on an approximate description of the wave field. The approximation significantly reduces the computational load involved in the reconstruction process and is based on an extension of the Rayleigh-Gans (Born) approximation. Quantitative results are presented of applying algorithms, based on the three techniques, to computer generated data.

An original technique for reconstructing binary images from as few projections as possible is introduced in Chapter 6. It is shown that only two projections are needed for reconstructing a convex cross-section of a binary object. The effects on the technique's performance of various practical considerations are analysed and quantitatively assessed. It is shown how the technique can be extended to non-convex objects. An original reconstruction algorithm is developed and is quantitatively illustrated.

Chapter 7 presents an original approach to reconstituting a three-dimensional object's shape and orientation from a set of its silhouettes, the relative orientations of which are unknown *a priori*. Because of the obvious applications of this in robotics, a review of robotic vision is presented. The underlying theory of this new approach and an algorithm developed from it are quantitatively illustrated.

Finally, Chapter 8 presents conclusions on, and suggests lines of further research into, the original methods introduced in this thesis.

During the course of my studies the following papers have been either published, or submitted for publication, or presented at conferences:

[1] R.D. Murch, D.G.H. Tan and D.J.N. Wall, Newton-Kantorovich method applied to two-dimensional inverse scattering for an exterior Helmholtz Problem, *Inverse Problems*, Vol 4, 1117-1128, 1988.

[2] R.D. Murch, B.K. Quek and B.M. McGregor, Determination of an object's orientation from silhouettes, *Proceedings of the New Zealand National Electronics Conference*, Christchurch, Vol 25, 59-64, 31 August - 2 September, 1988.

[3] D.G.H. Tan, R. D. Murch and R.H.T Bates, Algorithmic implementation of a global solution to the scalar inverse scattering problem, *Inverse Problems*, Vol 4, 1129-1142, 1988.

[4] D. G. H. Tan, R. D. Murch, and R. H. T. Bates, Inverse Scattering for Penetrable Obstacles, *Proceedings of the 1989 URSI International Symposium on Electromagnetic Theory* (Royal Institute of Technology, Stockholm, Sweden, August 14-17), 160-162, 1989.

[5] R. H. T. Bates, R. D. Murch and V.A. Smith, Manageable Multidimensional Inverse Scattering Theory, prepared by invitation for *Physics Reports*, 1990.

[6] R.D. Murch and R.H.T. Bates, Image Reconstruction from Projections IX: Binary Images, *Optik*, In preparation.

[7] R.D. Murch and R.H.T. Bates, Image Reconstruction from Projections XI: Reconstituting Object Shape and Orientation, *Optik*, In preparation.

Chapter 1

Introduction

A good beginning to the study of inverse problems is appreciation of how inverse methods are presently invoked in practice. The knowledge gained thereby highlights the assumptions that have been made to solve practical inverse problems. It also permits one to understand how practical constraints restrict the acquisition of ideal data. This helps to define the directions in which research needs to progress for improving inverse methods. So as to provide specific illustration of this viewpoint, §1.1 outlines the application of inverse theory to seismic methods in geophysics, ultrasonic methods in medical imaging and computed tomography. The experimental arrangements, types of display and also the methods by which observed data are typically processed are discussed. §1.2 suggests desirable improvements to the techniques described in §1.1, thereby identifying particular conventional assumptions which might benefit from appropriate modification.

Once specific applications of inverse theory have been introduced, it is appropriate to give a formal definition of the inverse problem, which is done, with the establishment of necessary terminology, in §1.3. This leads, in §1.4, to identifying preliminary tasks which need to be elucidated before the inverse problems of particular concern in this thesis can be addressed. Also summarised in §1.4 are established approaches to the solution of inverse problems, and the tendency of the solutions to be ill-posed. §1.5 discusses the impact of the dimensionality of space on inverse problems.

All theoretical developments are presented in three dimensions. For computational economy and expositional convenience, however, all quantitative illustrative, examples are two-dimensional.

1.1 Review of Practical Inverse Methods

Many solutions of the inverse problem are in practical use today. The solutions provide information which is valuable to many disciplines. There is always a demand, however, for improved inverse methods which provide more precise information. Before contemplating improving these inverse methods it is necessary to understand their status. This section reviews seismic methods in geophysics (§1.1.1), ultrasonic methods in medical imaging (§1.1.2) and computed tomography (§1.1.3).

1.1.1 Seismic Methods in Geophysics

Geophysics is the study of the physical nature and properties of the earth (Parasnis, 1979). A large part of this subject is concerned with the constitution of the earth's crust. Results of the labours of geophysicists are often used in oil prospecting, the location of water bearing strata, mineral exploration, highway construction and civil engineering. Because observation of the earth within its crust must be made indirectly, it is appropriate to invoke the theory of inverse problems when attempting to image the crust's interior. Details of the various conventional inverse methods appropriate in particular circumstances depend upon the particular emanations which are employed. Some commonly used emanations, other than seismic waves, are magnetic, electromagnetic and gravitational fields, electric currents, radioactivity and induced polarisation (Parasnis, 1979). Because seismic methods have had by far the greatest economic impact of all these emanations, they are concentrated on here.

Seismic methods rely on elastic vibrations or waves, travelling at different velocities in different materials (Sengbush, 1983). The principle is to generate such waves at a particular location and determine at a number of other locations the times of arrival of the elastic vibrations reflected from interfaces between different rock formations, enabling the shapes and positions of the interfaces to be estimated.

The standard method of generating seismic waves is by explosive charges set in "shot holes" at depths of up to a few hundred metres in the Earth. Other methods such as dropping heavy weights onto the Earth from a large height are also used (Parasnis, 1979). The seismic waves are detected with geophones, the most common type (the electromagnetic geophone) is based on the principle that moving a coil relative to a magnetic field induces an electric voltage in the coil.

When a geological area is to be probed by seismic imaging, the shot holes and geophones must be appropriately located. They are commonly laid out as a linear array. The shot holes are usually equi-spaced along a particular line while the geophones are planted firmly on the ground on either side of the holes. Typical positions of the geophones with respect to a shot hole are shown in fig. 1.1a. After each shot the geophones are repositioned around the next shot in the same way. Recording equipment is connected to the geophones so that the data collected after each shot can be processed later.

A popular method of processing the recorded data is known as the common depth point stack (cdp) (Schoenberger, 1984). The main principle of this method is to associate together all signals which are reflected from the same point. This allows all the signals reflected from a point to be processed together. An improved reconstruction of the location of, and intensity of the signal reflected from, each point can then be obtained. To this end the observation structure shown in fig. 1.1a is reorganised into the virtual structure typified by fig. 1.1b. All the recorded signals can then be considered to be due to reflections from points along a line vertically below the centre of the structure. Geophones further from the structure's centre receive reflected signals which are, however, delayed because these signals travel a greater distance through the Earth. To compensate for these longer delays, recordings from the geophones are time-shifted so that signals reflected from the same point all appear to occur at the same time. This is known as normal move out correction. Once the normal move out correction is made all the geophone recordings are added together. This results in an improved signal

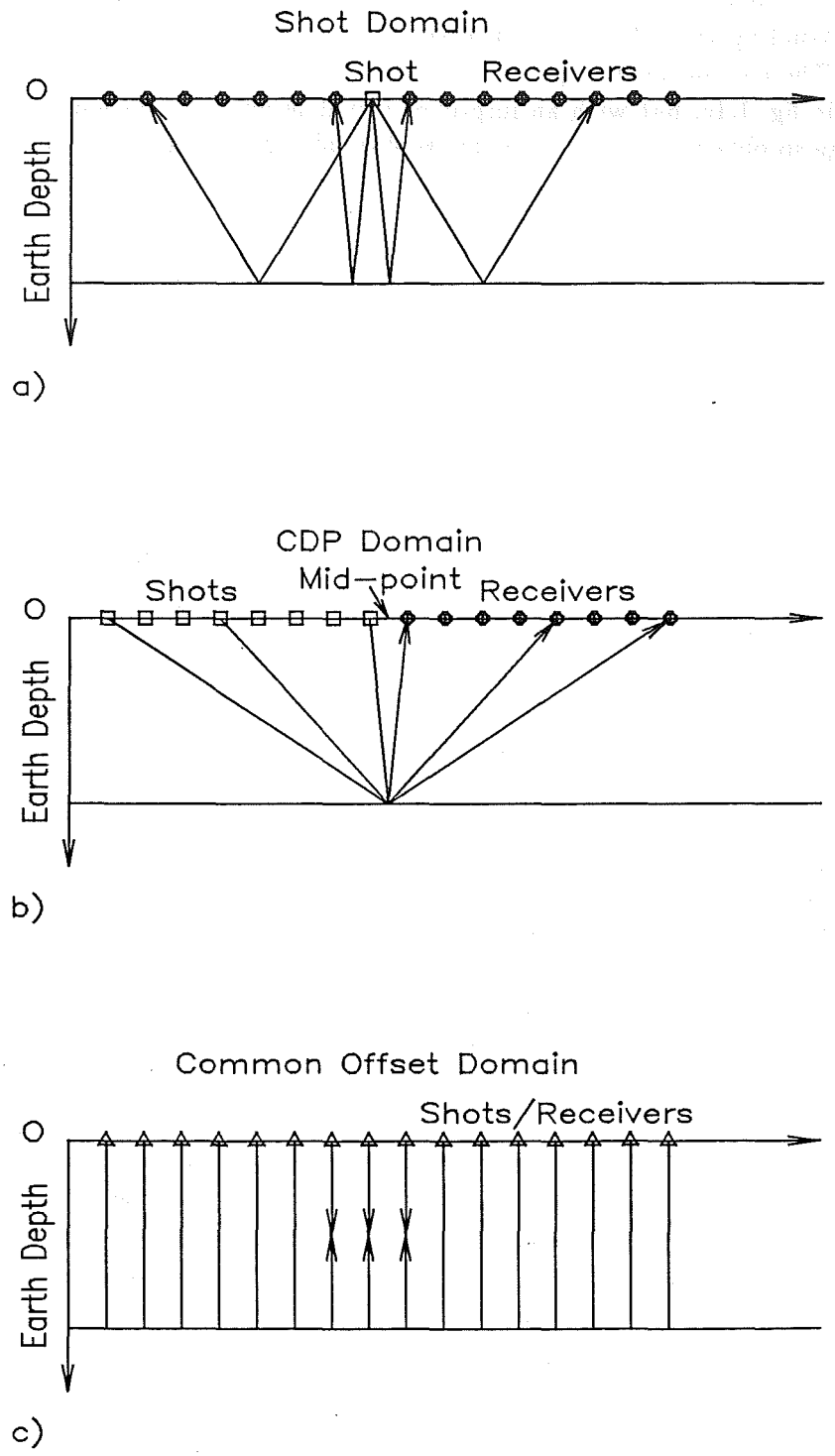


Figure 1.1: Observation structure for seismic imaging. a) The actual observation structure used in practice. b) The virtual structure used in common depth point stacking. c) Equivalent structure of common depth point stacking.

to noise ratio for the resultant signals reflected from each point. The resultant signal represents a reconstruction of the reflectivity of the Earth along a vertical line below the centre of the observation structure. An image is built up of the Earth's reflectivity by reconstructing the reflectivity at adjacent vertical lines and mapping them appropriately. The common depth point method is equivalent to the observation structure depicted in fig. 1.1c, but with an improved signal to noise ratio (Schoenberger, 1984). The image so obtained can be used for geological exploration. However, a number of shortcomings of the cdp process have been recognised since it was originally introduced over 30 years ago. It is now common practice to carry out additional processing, after cdp stacking, to correct for some of the known shortcomings. The additional processing which is involved is discussed next.

The velocity of seismic waves near the surface of the Earth varies from point to point. Unless properly accounted for, these variations affect the quality of the reconstructed image (Schoenberger, 1984). Because the variations can usually be inferred from geological bore hole studies (Schoenberger, 1984), their detrimental effect can be corrected. The variation in seismic velocity near the Earth's surface correspondingly alters the delay each reflected signal experiences before reaching a geophone. The variations in delay cause the images of the reflecting points to be distorted. The distortions are corrected by time-shifting each recording by an amount governed by the known seismic surface velocity variations. These corrections are usually made when normal move out correction (introduced in the previous paragraph) is performed because both processes carry out time-shifting of the recorded signals. The resultant image represents an improvement because the reflecting points are more accurately positioned.

In the common point stack method it is assumed that the reflected signals received by each geophone are predominately from a single direction. Frequently, in fact, the seismic signal recorded by a single geophone is a superposition of seismic waves of comparable amplitude originating from a range of directions. A better description of the seismic data acquisition process is that of a recording of upward travelling waves emanating from points in the Earth. To produce an improved image using this description the recordings in the cdp process are used to construct a field satisfying the wave equation (Sengbush, 1983). As a result, the recorded waves are propagated backward or migrated from the surface of the Earth to the reflector locations to produce a more faithful image. However the velocity of the seismic waves in the Earth must be known *a priori* or assumed. This process has become known as migration and is usually performed after normal moveout correction in cdp processing (Berg and Woolverton, 1986).

Distortion of the image also occurs because the explosive shot does not produce an ideally sharp source pulse. The actual source pulse can be modelled as a convolution of an ideally sharp pulse with a known distorting function which is determined from the explosive type, shot position and the seismic recording itself (McQuillin *et al.*, 1984). Standard deconvolution techniques can be invoked for final enhancement of the image by reducing the effects of the distorting function (Sengbush, 1983).

1.1.2 Ultrasonic Methods in Medical Imaging

The development of imaging by ultrasound began with the invention of sonar in 1917 (Wells, 1977). Since then imaging by ultrasound has been employed in a large number of disciplines. One important application of ultrasonic imaging has been for medical

diagnosis of, for instance, the brain, eye, heart and abdomen and also in the assessment of foetal development during pregnancy (Wells, 1977). The growth in the use of ultrasound for medical diagnosis can be partly attributed to the current evidence that it is safe as compared to the known hazards of ionising radiation (Stratmeyer and Lizzi, 1986). However constant vigilance over its safety is warranted as our knowledge of thermal effects, cavitation and acoustic streaming and other biological phenomena increases. At high powers ultrasound has been shown to cause permanent changes to human tissue. These effects are harnessed, for example, in various surgical operations and physiotherapy (Wells, 1977). The most widely used ultrasonic imaging technique at present is B-scan which is able to provide a 2D image (Fatemi and Waag, 1983). This section describes the principles of the B-scan modality.

B-scan imaging relies on ultrasound being reflected from organs inside the body. The strengths and delays of the reflections enables the structures and positions of the body organs to be estimated (Halliwell, 1987).

The ultrasound for a B-scan medical instrument is normally generated by a piezo-electric transducer. The transducer is shaped so that it emits a narrow beam of ultrasound. The ultrasonic frequencies are generally set in the 1 to 10 MHz range, which is chosen as a compromise between high resolution for the image and acceptable attenuation of the signal passing through body tissue (Taylor, 1979). The ultrasonic transducer also serves conveniently as a receiver because the piezo-electric effect is reciprocal.

The positioning of the transducer, on the surface of the body, for B-scan imaging is shown in fig. 1.2a. A short pulse is transmitted from the transducer. The reflected signals caused by interaction of the ultrasound with the body are recorded by the same transducer. Because the transmitted pulse is emitted in a narrow beam, the observed signal characterises the reflectivity of the body along a small tube originating from the transducer (Wells, 1977).

The reflected signals are conveniently displayed on a cathode ray tube. The signals are electronically mapped onto a line on the cathode ray tube which represents the reflectivity along the narrow beam of ultrasound emitted by the transducer (see fig. 1.2b). A complete 2D image of the body is built up by successively changing the direction of the transmitted beam and mapping the received signals onto corresponding lines on the cathode ray tube. Changing the direction of the transmitted beam can be effected mechanically but it is nowadays usually done electronically by using a phased array transducer (Taylor, 1979).

The interpretation of images produced by B-scan is difficult for untrained personnel. However, trained personnel are able to make accurate diagnoses from the images (Kim *et al.*, 1987).

1.1.3 Computed Tomography

Computed Tomography (CT) has developed explosively since Hounsfield (1972) first demonstrated a practical CT scanner (Robb, 1985). Since then the most prolific application has been in clinical medicine, enabling physicians to examine internal structures of the body in fine detail prior to surgery. The revolution that CT has caused in medical imaging is indicated by the awarding of the Nobel Prize in medicine in 1979 to Allan Macleod Cormack and Godfrey Newbold Hounsfield for their independent contributions to the field (Robb, 1985). During the 1980s, CT machines have been installed in

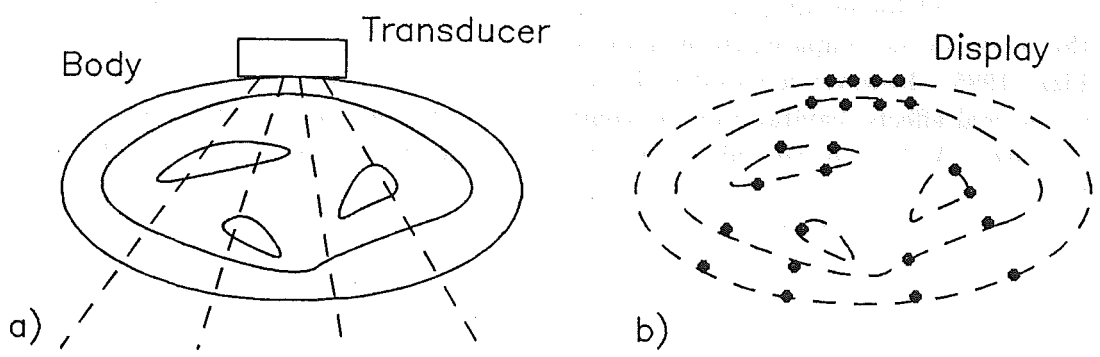


Figure 1.2: Operation of ultrasonic B-scan. a) The ultrasonic transducer is placed on the human body and the ultrasonic beam (dashed line) is electronically moved over a range of directions. b) An image of the reflectivity of the human body is displayed on a cathode ray tube. The solid dots correspond to points in the body which have caused the ultrasonic beam to be reflected.

tions to the field (Robb, 1985). During the 1980s, CT machines have been installed in a great many clinics and hospitals around the world (Herman, Ed.) (1983).

CT relies on tissues and tumours of different types attenuating x-rays differently. The attenuation an x-ray experiences while it passes through the body equals the integral of the body's attenuation coefficient along the line the x-ray travels. In practice these integrals can be acquired by detecting the strength of x-rays, emitted by a source, which have transversed the human body. By collecting integrals through a cross-section of the body over a range of angles co-planar to the cross-section an image of the attenuation coefficient at each point in the cross-section can be obtained (Radon, 1917,Bates and McDonnell, 1989, Chapter 5). Because the attenuation of most materials is closely proportional to their materials densities the image of the cross-section's attenuation coefficients is effectively an image of the density (Bates and McDonnell, 1989, Chapter 5). Through this method CT is able to provide accurate quantitative reconstructions of densities of the cross-sections of bodies.

Commercial x-ray CT scanners generally comprise six modular subsystems (Robb, 1985). These are the gantry which supports the x-ray source and detectors, the moveable table on which the patient lies, the x-ray source, the array of detectors, the associated computer and the image display system.

The x-ray source is usually an air-cooled rotating anode pulsed tube. It is polyenergetic with an average energy of about 50 keV. Because the standard reconstruction techniques implicitly assume that the beam is mono-energetic, the polyenergetic nature of the beam can cause a type of image degradation known as beam hardening (Alvarez and Macovski, 1976; Garden, 1984). An x-ray beam collimator at the source limits the conical beam to a fan beam, thereby eliminating x-ray dosage which does not

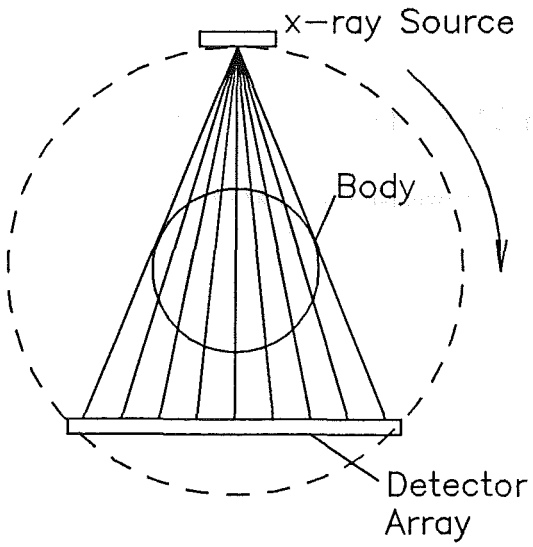


Figure 1.3: Basic fan beam geometry used in modern CT scanners. A single x-ray source produces a fan beam of x-rays sensed by an array of detectors. Simultaneous rotation of the x-ray source and detector array through 360° enable integrals of the attenuation coefficient through the body to be collected.

contribute to the reconstructed image. The design of the x-ray detectors is critical for the resolution and contrast, set by the sizes and efficiencies of detectors respectively, of images formed by CT scanners. Three types of detector which are commonly used are those termed scintillation, gas and solid state (Robb, 1985). Aperture diameters are between 0.75mm and 2.0 mm. The gantry provides rotational movement and supports the source and detectors. It is fabricated to strict mechanical tolerances to ensure precise alignment of the x-ray source and detectors relative to the body being imaged. Most modern CT gantries employ what is known as fan beam scanning, as illustrated in fig. 1.3. Simultaneous rotation of the x-ray source and detector array through 360° enables a body's cross-section to be illuminated from all directions in the plane of the cross section. The data gathered for each direction of illumination corresponds to integrals of the body's cross-section attenuation along the paths of the x-rays. Fan beam scanning of a single cross section of a body can typically be completed in under five seconds.

Once the scan is completed the collected data must be processed to form an image. The data are treated as projections through the cross-section of the body so that algorithms based on the projection theorem can be invoked to generate an image of the cross-section (Bates and McDonnell, 1989). This involves converting around one million observations into a 256x256 or 512x512 pixel image. Today's minicomputers are able to perform this processing within a few seconds. Once the image has been computed it must be displayed. The display device is normally a cathode ray tube capable of 512 grey levels, or colour with 1024x1024 pixel resolution. Image manipulations, such as emphasising certain regions and producing histograms, are also possible and are now incorporated into commercially available display modules.

The clinical diagnostic applications of CT have now become routinely accepted,

can be recognised by the untrained eye. It is this image quality which is characteristic of CT scanners (Robb, 1985).

1.2 Research Directions for Practical Inverse Methods

The desire to obtain more precise information about remote objects is seemingly insatiable. This section discusses some of the research directions being followed with the aim of improving the inverse methods which are presently invoked in practice.

1.2.1 Seismic Methods in Geophysics

One motivation for improving seismic imaging techniques is the increasing demand for oil world wide and the decreasing number of deposits available. More accurate techniques are required to discover whether oil deposits are likely to exist in terrain which is difficult to image and to form new images of regions which have been prospected with the aid of older techniques.

Historically, the first imaging techniques were based on assuming the earth could be modelled as a series of spherically symmetric layers. Earth structure departing radically from this model could not be imaged. Today, methods based on so-called migration techniques (see §1.1.1) can image asymmetrical or amorphous regions of the Earth, provided the local seismic velocities can be reasonably accurately obtained (Sengbush, 1983). If the estimated velocities are significantly erroneous, apparent positions of reflecting boundaries are seriously misplaced.

Another drawback of migration techniques is that the only property of the Earth which can be imaged is reflectivity. Specific rock types and densities can only be identified after the images are interpreted in the light of geological expertise (Daily, 1986).

Direct recovery of acoustic or elastic parameters of the Earth, with little reliance on *a priori* information concerning the spatial variation of seismic velocity, is the ultimate aim of all current research into seismic techniques (Lines, 1986). The approaches presently being investigated are perhaps better described as seismic inversion rather than seismic imaging. Such inversion techniques are based on full-wave descriptions of seismic propagation. However, they need considerable further development in order for them to be of routine practical use (Daily, 1986; Lines, 1986). The methods are very expensive computationally and more study is needed of their sensitivity to the corrupting effects of noise. Special geometries of the Earth structure must sometimes be assumed in order to generate apparently unique images. The majority of these methods are also predicated on seismic sources which give rise to plane waves, something which rarely occurs in practice. It is only when these shortcomings are overcome that it will be possible to claim that valid inversion techniques have been developed.

Because of the difficulties associated with seismic inversion research is also being performed on inversion methods which incorporate adequate descriptions of wave motion (not necessarily full wave descriptions) and which are also manageable in practice. Some of the research being pursued to this end has been reported by Lytle and Dines (1980), Clayton and Stolt (1981), Beylkin (1985), Devaney (1987) and Lo *et al.* (1988).

1.2.2 Ultrasonic Methods in Medical Imaging

Ultrasonic imaging has become the dominant imaging modality in many areas of medical diagnosis. However, comparison of ultrasound B-scan with x-ray CT images, for example, shows that the information displayed in B-scan images is much less quantitative. Although the initial development of ultrasonic B-scan was rapid, it has now reached a plateau without any significant recent technological advances having been reported (Moss, 1982).

Spatial resolution of B-scan is comparatively poor and the images contain speckle (Moss, 1982; Shankar, 1986). Quantitative interpretation of B-scan images, such as unambiguous characterisation of different types of tissue, is heavily dependent on the experience operators gain by experiment (Greenleaf, 1983). There is also a need for improved understanding of transducer operation so that specialised transducers can be designed for certain surgical operations (Moss, 1982).

Improvements in transducer design are allowing greater resolution by providing smaller beamwidths and narrower pulses. However, mere reliance on superior hardware must limit possible improvements in image quality (Fatemi and Waag, 1983; O'Donnell, 1988). This is because ultrasonic propagation can rarely be described, even approximately, in terms of straight rays (such as is permissible, in almost all instances, for radar). Phenomena such as diffraction and refraction cause the ultrasonic rays to deviate significantly from being straight. This suggests that imaging methods like B-scan, which are based on a straight ray model of ultrasonic propagation, can rarely be expected to produce faithful images. Substantial improvements in image quality will only be possible after a model of propagation which incorporates phenomena such as refraction and diffraction has been devised. When (and if) this is done, it should enable images to be formed which quantitatively relate to some parameter characterising a physical property of the body (Greenleaf, 1983).

Ultrasonic imaging methods which are based on a full-wave description of propagation have been developed by Roger (1981), Johnson and Tracy (1983), Colton (1984), Kristensson and Vogel (1986) and Colton and Monk (1988). They suffer, however, from certain drawbacks which prevent them being used in practice. Some of them demand large computational power whilst others can only accurately image a restricted set of objects. As in seismic imaging, there is a need for techniques which can combine an adequate description of wave motion with practical exigencies. Relevant studies are presented by Bates *et al.* (1976), Beylkin (1985a), Dines and Goss (1987), Devaney (1987), Schultz and Jaggard (1987).

1.2.3 Computed Tomography

CT has become the standard to which a number of other imaging techniques are compared. However, there are certain aspects of CT images which medical practitioners would like to see improved.

CT suffers from beam hardening, as discussed in §1.1.3. This causes an apparent decrease in the attenuation coefficient as the penetration of the beam increases (Alvarez and Macovski, 1976; Garden, 1984). The most objectionable result is the generation of "streaky" artefacts which appear (in the reconstructed images) to emanate from dense objects (Bates *et al.*, 1983). It has also become desirable to generate images quickly (within one second, say) so that moving organs can be viewed in real time (Moss, 1982).

The problems mentioned above are mainly due to comparatively large detector beamwidths, non-monochromatic x-ray sources and relatively slow scanning times, which can, in principle, be remedied by improved hardware. The fundamental model upon which CT is based closely approximates the basic physics. This means that improvements in the imaging hardware, rather than the propagation model, are more likely to result in improvements in image quality. This contrasts markedly with many other inverse methods which require improvements to their models of wave propagation.

1.3 Archetypal Problems

Having descriptively examined, in §1.1, some inverse problems which occur in practice, it is now appropriate to formally define a general, or archetypal, inverse problem (see §1.3.3). Inverse problems can only be systematically analysed if they are presented within a general mathematical framework. Mathematical analysis of such problems allows reliable solutions to be obtained, where the term reliable here implies that the levels of validity of the solutions, and the conditions under which they can be invoked, can be quantitatively characterised. Since inverse problems can only be understood in the contexts of their associated direct problems (recall that both inverse and direct problems have been discussed in general terms in the preface) an archetypal direct problem is also formulated in this section (see §1.3.2). Terminology and notation suitable for such a framework are introduced in §1.3.1.

1.3.1 Terminology and Notation

It is convenient to partition K-dimensional space as indicated in fig. 1.4. Although this thesis is only concerned in detail with K=3 and K=2, it seems appropriate to start out in this section by establishing notation for a space of arbitrary dimensionality. The whole of K-dimensional space is denoted by Υ, an arbitrary point in which is labelled by the position vector $\mathbf{x}$ with respect to an arbitrarily chosen coordinate origin O. The scattering region (taken to be finite) is denoted by Υ_-, the surface of which denoted by σ. The hyper-spherical surface σ_- centred on O, inscribing Υ_-, partitions Υ_- into Υ_{--} and Υ_{-+} interior and exterior, respectively, to σ_-. The part of Υ exterior to σ is denoted by Υ_+. The hyper-spherical surface σ_+ centred on O, circumscribing Υ_-, partitions Υ_+ into Υ_{+-} and Υ_{++} interior and exterior, respectively, to σ_+. Invoking set theoretical notation, the above definitions are formalised by

$$\Upsilon = \Upsilon_- \cup \sigma \cup \Upsilon_+ \quad \text{and} \quad \Upsilon_- \cap \sigma = \Upsilon_+ \cap \sigma = \emptyset \tag{1.1}$$

where $\emptyset$ is the empty set, with

$$\Upsilon_- = \Upsilon_{--} \cup \sigma_- \cup \Upsilon_{-+} \quad \text{and} \quad \Upsilon_{--} \cap \sigma_- = \Upsilon_{-+} \cap \sigma_- = \emptyset \ , \tag{1.2}$$

and

$$\Upsilon_+ = \Upsilon_{+-} \cup \sigma_+ \cup \Upsilon_{++} \quad \text{and} \quad \Upsilon_{+-} \cap \sigma_+ = \Upsilon_{++} \cap \sigma_+ = \emptyset \tag{1.3}$$

The generalised constitutive parameter χ characterises properties of interest within the scattering region Υ_-. The physical process used to probe or sense χ within Υ_- is referred to as the emanations, which are characterised by the wave function ψ. In general, ψ can be partitioned into a probing emanation ψ_i and a scattered emanation

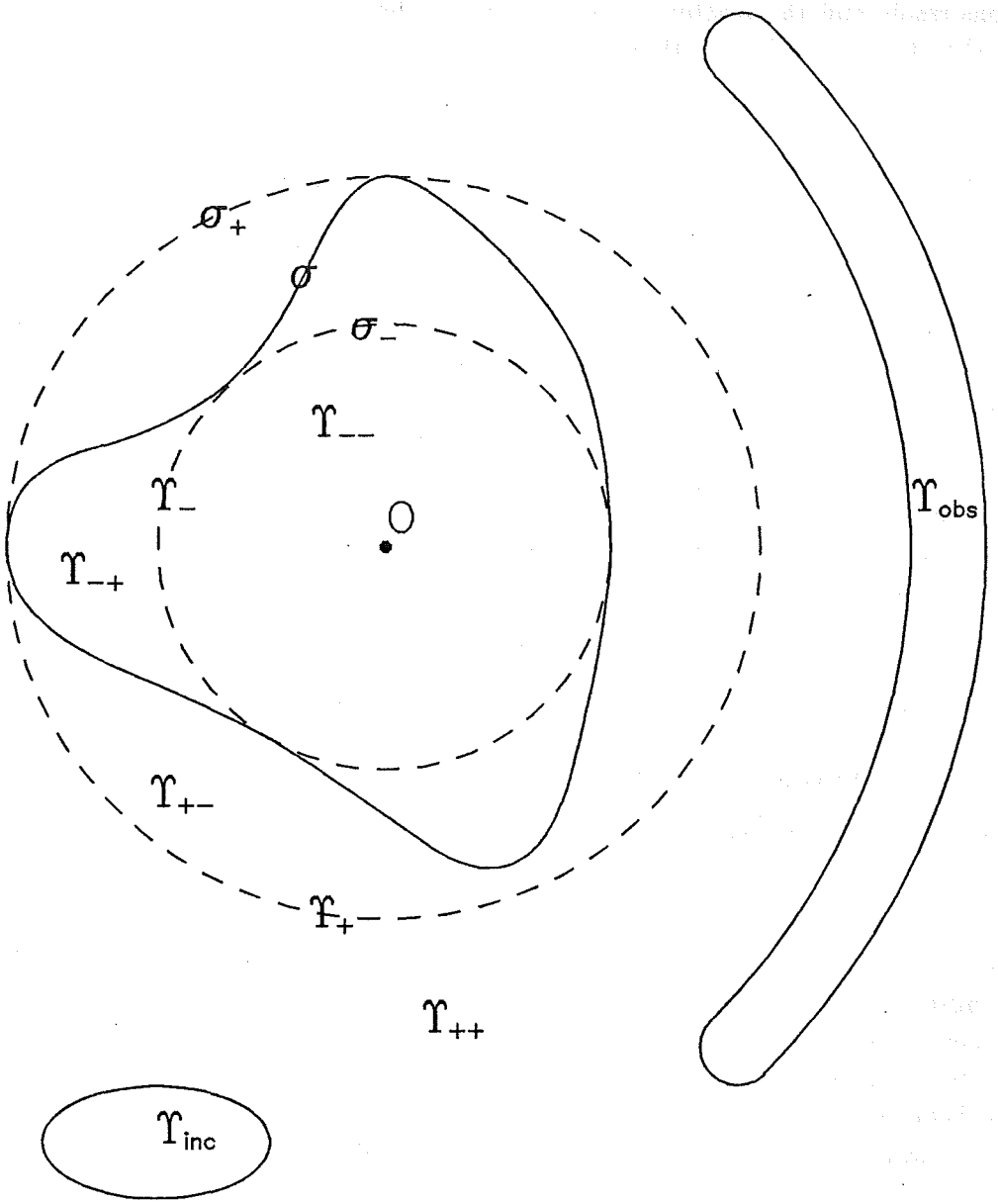


Figure 1.4: Partitioning of space. The symbols Υ and σ refer to hyper-volumes and hyper-surfaces in K-dimensional space.

ψ_s. The latter emanation results from the probing emanation interacting with the generalised constitutive parameter. Therefore

$$\psi = \psi_i + \psi_s \tag{1.4}$$

The regions Υ_{inc} and Υ_{obs} are where, respectively, the sources of the incident emanations reside and the scattered emanations are observed. Both of these regions are defined to be outside Υ_-, so that

$$\Upsilon_{inc} \subset \Upsilon_{++} \quad \text{and} \quad \Upsilon_{obs} \subset \Upsilon_{++} \tag{1.5}$$

It is emphasised that both two-dimensional (K=2) and three-dimensional (K=3) scattering can be conveniently described with the aid of the above notation. In general, when K=3, Υ_- is a conventional volume of arbitrary shape, with σ_- and σ_+ being conventional spherical surfaces. Now consider the special situation of Υ_- being an infinite cylinder of arbitrary cross-section (whose shape is defined by σ which in this instance is of course a cylindrical surface). If, in addition, ψ_i exhibits no variations in the direction of the cylinder's axis (e.g. it is a plane or cylindrical wave oriented appropriately to this axis) then ψ only varies in directions lying in planes perpendicular to the said axis, implying that the scattering is effectively two-dimensional.

When the scattering is effectively two-dimensional, as indicated in the previous paragraph, it can be treated mathematically as actually two-dimensional (i.e K=2). Then σ reduces to an arbitrary closed curve, with σ_- and σ_+ being perimeters of circles centred on the origin O of coordinates. In later Chapters it is often found convenient (when K can be taken to be 2) to accord to $\mathbf{x}$ the Cartesian coordinates (x, y) and the cylindrical polar coordinates $(\rho; \phi)$, or alternatively $(r; \theta)$.

1.3.2 The Direct Problem

The direct problem is here stated as

$$\psi_s = \Lambda(\chi, \psi_i) \tag{1.6}$$

where Λ is an operator which describes the interaction of the emanations with the generalised constitutive parameter χ. The direct problem can consequently be alternatively stated as: given the generalised constitutive parameter, calculate the result of interactions of the emanations with it.

This definition and the terminology introduced in §1.3.1 are adequate for specifying most direct problems of practical interest. For example, the associated direct problem for ultrasonic B-scan is to calculate ψ_s in Υ_{obs} which is the exterior surface of the ultrasonic transducer. The scattering region Υ_- is the body, the generalised constitutive parameter is the body's reflectivity, and the emanations are ultrasonic waves. The probing emanations originate from the ultrasonic transducer, the exterior surface of which is also Υ_{inc}.

1.3.3 The Inverse Problem

The inverse problem is the mathematical inverse of (1.6) and is stated here as

$$\chi = \Lambda^{-1}(\psi_s, \psi_i) \tag{1.7}$$

where Λ^{-1} is the mathematical inverse of the operator Λ. It should be noted that the domain of χ is within Υ_-, whereas ψ can only be observed in practice within Υ_{obs}. The inverse problem can consequently alternatively be stated as: given the emanations within Υ_{obs}, calculate the distribution of the generalised constitutive parameter throughout Υ_-.

This definition and the terminology introduced in §1.3.1 are adequate for specifying many inverse problems of practical interest. For example, when inverting seismic data (i.e. forming seismic images) the scattering region Υ_- is the Earth, the generalised constitutive parameter is reflectivity, and the emanations are seismic waves. The region Υ_{inc} is where the apparatus which generates the shots is. As explained in §1.1.1, these shots are the source of the probing emanations. The scattered emanations are the reflected signals which are observed at the geophones in the region Υ_{obs}. Seismic imaging thus involves determining the reflectivity of the Earth from seismic observations made outside the Earth.

All this can be summarised. The direct problem can be considered as, given a cause find the effect, whilst the inverse problem is, given an effect find the cause.

1.4 Solving the Inverse Problem

The inverse problem is mathematically defined in (1.7). The solution to (1.7) in specific situations represents mathematical descriptions of imaging methods. To find a solution to (1.7), the operator Λ^{-1} must first be specified. The form of Λ^{-1} depends on the underlying physics of the interaction of the emanations with the generalised constitutive parameter χ. In most situations, χ can only be expressed as an implicit non-linear function of ψ, mainly because ψ_s can only be observed outside the body whereas χ exists entirely inside the body. Solutions deduced from Λ^{-1} can usually be expected to be either non-unique, non-existent, or to depend non-continuously on the data. An operator characterised by one or more of these properties is ill-posed (Baltes, 1980; Sabatier, 1983). An ill-conditioned system is here defined to be a system that manifests itself numerically as the consequence of either an ill-posed operator or an operator which is nearly so.

Ill-posedness can usually be ameliorated by regularisation which can be regarded as a technique for incorporating *a priori* information to restrict the set of possible solutions (Baltes, 1980; Sabatier, 1983). There are two commonly adopted approaches which overcome the inherent implicit and non-linear nature of Λ^{-1}. One either introduces approximations to reduce Λ^{-1} to a explicit linear expression or one invokes inversion theory (Sabatier, 1983). These two approaches are discussed next.

Approximations can be invoked which enable the operator Λ^{-1} to be expressed such that it depends explicitly and linearily on the scattered field. A solution to the inverse problem is then obtained by directly evaluating the operator. The structure of the corresponding direct problem is usually invoked to develop an explicit linear expression for Λ^{-1}. This is effected by approximating the direct problem to obtain an explicit linear expression for Λ. This expression can then be inverted by some standard linear technique to obtain, in accord with (1.7) and (1.6), an explicit linear expression for the inverse problem. However, a solution can only be considered satisfactory if the approximations are deemed acceptable. In many situations, the structure of the operator Λ^{-1} is so intricate that it cannot be approximated (or rather, no satisfactory

approximation has yet been found) without generating wildly inaccurate solutions. For example, seismic and ultrasonic imaging techniques sometimes produce severely distorted images because the operators involved are being approximated poorly. In the case of CT, however, only small approximations to the operator are required with the result that the images produced are relatively undistorted.

The other approach which overcomes the implicit non-linear nature of Λ^{-1} is to invoke what is called inversion theory (Rall, 1969; Sabatier, 1983). This approach is based on the notion that Λ which is the operator for the direct problem, maps the generalised parameter χ to the scattered field ψ_s. This suggests that a model of the unknown χ can be guessed, so that the corresponding field ψ_s can be calculated. If the calculated scattered field is sufficiently close to the observed ψ_s then χ can be said to be a solution to the inverse problem. If ψ_s is not sufficiently close to the observed scattered field then χ must be altered until ψ_s becomes sufficiently close. This idea leads to a general iterative solution to inverse problems which hopefully locates correct solutions. A thorough understanding, however, of the underlying physics is necessary to ensure satisfactory solutions are obtained (Baltes, 1980; Sabatier, 1983)

Because, in both of these approaches, the direct problem plays a central part in the solution of the inverse problem, it is important to understand the analytical niceties of the direct problem. Accordingly, Chapter 3 is devoted to the direct problem.

1.5 Impact of Dimensionality of Space on Inverse Problems

There is a considerable literature on inverse scattering problems posed in one spatial dimension. Some key references are Gelfand and Leviatan (1955), Chadan and Sabatier (1977) Newton (1981) Hashaby and Mittra (1987) and Bruckstein and Kailath (1987).

Study of the references quoted above reveals that one-dimensional inverse scattering theory is by now comparatively complete and very well understood, which contrasts spectacularly with the state of the general theory for more than one spatial dimension. There are two reasons for this (Bates *et al.*, 1990).

There is a physical intuition concerning the interaction of linear wave motion with scatterers, which is so ingrained in radio/radar engineers, acousticians and quantum physicists, that it must essentially characterise natural wave phenomena in the great majority of physically interesting situations. This intuition claims that wave scattering can be represented in terms of rays. That is, wave motion penetrates a body in such a way that the energy propagates along curves. In one dimension then, one knows *a priori* the directions of all rays for a one-dimensional inverse scattering problem. There is only one direction (i.e. that corresponding to the single spatial dimension) in which they can travel! In more than one dimension the rays travel along arbitrary curves which are not known *a priori.* Consequently, the inverse problem acquires an extra degree of intricacy in more than one dimension.

The second reason is more subtle and relates to it always being possible to transform a wave equation involving a spatially variable refractive index, but depending upon only a single spatial Cartesian coordinate, into a Schrödinger (or plasma wave) equation which is correspondingly easier to solve (Bates and Wall, 1976; Chadan and

Sabatier, 1977).

It can be immediately appreciated from the above that the technical (due to not being able to transform to the equivalent of the Schrödinger equation) and conceptual (associated with being unclear *a priori* as to the actual ray paths) difficulties experienced with three-dimensional inverse scattering problems are undiminished to any significant degree in two dimensions. On the other hand, the computational effort required to simulate two-dimensional inverse scattering scenarios is much less than for their three-dimensional manifestations. Therefore, preliminary studies of inversion techniques carried out in two dimensions can be exceedingly valuable, which is why the numerical simulations presented in this thesis are for two dimensions.

Chapter 2

Mathematical Preliminaries

This Chapter is concerned with the various mathematical techniques, together with their associated terminology and notation, which are needed for solving direct and inverse problems involving wave-like emanations. Actual solutions to these types of problem are presented in Chapters 3,4 and 5. These mathematical preliminaries provide the proper setting for the later Chapters and allow the solutions to be developed without unduly disrupting them to make every point precise.

An important preliminary is to find the form of the operator Λ describing the interactions between the wave-like emanations and the generalised constitutive parameter (see §1.4). In §2.1 mathematical models are developed for describing these interactions through differential equations. The applicability of the models is discussed and their formulation in the time and frequency domains are presented. The relative merits of the time and frequency domain formulations are also discussed. §2.2 presents various supplementary definitions and mathematical techniques which are useful in solving direct and inverse problems.

2.1 Mathematical Models of the Physics of Scattering

Wave motion in various media can be modelled by differential equations. The wave motion can be formulated in either the time domain or the frequency domain.

2.1.1 Time Domain Formulations

Under certain sets of conditions, applying in many situations of practical interest, electromagnetic, elastic and acoustic emanations can be described by a common mathematical model. These emanations can be adequately represented in the time domain by a complex valued scalar wave function $\Psi(\mathbf{x}, t)$, hereafter called the field, which exists at all points $\mathbf{x} \in \Upsilon$ in arbitrary dimensional space and all instants $t \in R$ in time (Felson and Marcuvitz, 1973, Chapter 1). Any individual Cartesian component of a vector field can also be described by $\Psi(\mathbf{x}, t)$ provided there is negligible cross coupling between the field's components. The Cartesian coordinate in the direction of the chosen component of the vector field is denoted by ξ. When the medium is source free, linear, time invariant and isotropic, the emanations can be described by the partial differential equation

$$\nabla^2\Psi(\mathbf{x}, t) - \beta(\mathbf{x})\dot{\Psi}(\mathbf{x}, t) - (\nu(\mathbf{x})/c)^2\ddot{\Psi}(\mathbf{x}, t) + \mu(\mathbf{x})\Psi(\mathbf{x}, t) = 0 \qquad (2.1)$$

where c is the free-space wave speed and $\beta(\mathbf{x})$, $\mu(\mathbf{x})$ and $\nu(\mathbf{x})$ are constitutive parameters which are related in table 2.1 to the properties of the medium for each type of emanation (Morse and Feshbach, 1953, Chapter 2; Born and Wolf, 1970, Chapter 1). Table 2.2 lists the symbols used in this thesis to represent properties of the media.

Emanation	$\beta(\mathbf{x})$	$\mu(\mathbf{x})$	$\nu(\mathbf{x})$
Electromagnetic	$\mu_0\sigma$	0	$\sqrt{\frac{\epsilon}{\epsilon_0}}$
Elastic	0	$\frac{2\rho\nabla^2\rho-3\nabla\rho\cdot\nabla\rho}{4\rho^2}$	$\sqrt{\frac{\rho}{\lambda}}$
Acoustic	0	$\frac{2\rho\nabla^2\rho-3\nabla\rho\cdot\nabla\rho}{4\rho^2}$	$\sqrt{\frac{\rho}{\kappa}}$

Table 2.1: Constitutive parameters $\beta(\mathbf{x})$, $\mu(\mathbf{x})$ and $\nu(\mathbf{x})$ related to the properties (themselves listed in Table 2.2) of propagating media for electromagnetic, elastic and acoustic wave motion.

Property	Symbol
Free-space permeability	μ_0
Permittivity	ϵ
Free-space permittivity	ϵ_0
Electrical conductivity	σ
Density	ρ
Compression modulus	κ
Bulk modulus	λ

Table 2.2: Properties of media and the respective symbols used to represent them.

To derive the relationships between the properties of the medium and the constitutive parameters, certain limitations on the medium must be imposed for each type of emanation. These limitations imply that (2.1) is only valid under restricted sets of conditions. These conditions are, however, of considerable practical interest.

To allow electromagnetic emanations to be described by (2.1), $\Psi(\mathbf{x},t)$ is taken to represent a single component of the electric or magnetic field. Because the great majority of media are non-magnetic, negligible error results in general from assuming the permeability of the medium to be constant and equal to its free-space value μ_0. A more significant limitation is that the permittivity ϵ cannot be allowed to exhibit appreciable spatial variations in the ξ-direction. The implication here is that $\nabla(\Psi(\mathbf{x},t)\frac{\partial}{\partial\xi}\log_e\epsilon$ must be everywhere small compared with both $\nabla^2\Psi(\mathbf{x},t)$ and $\frac{1}{c^2}\ddot{\Psi}(\mathbf{x},t)$ (Jones, 1964, Chapter 1; Born and Wolf, 1970, Chapter 1). While the class of media that satisfies this limitation is palpably restricted, it is wide enough to be of considerable practical interest.

Elastic emanations can be modelled by (2.1) if $\Psi(\mathbf{x},t)$ represents the displacement of the particles in the medium. Elastic emanations occur in two forms, known as primary or pressure p-waves and secondary or shear s-waves (Sengbush, 1983). The only limitation on the medium for p-waves is that there should be negligible cross-coupling

to s-waves. The restriction on s-waves is much more severe because the medium must belong to the class for which there is negligible shear in the ξ-direction (Spencer, 1980). It is thus seen that the limitations on the allowable classes of media are similar for electromagnetic waves and elastic shear waves.

Acoustic emanations can also be described by (2.1) if $\Psi(\mathbf{x}, t)$ represents the velocity potential for the particles in the medium (Wilcox, 1984, Chapter 1). Since acoustic emanations can also occur in p and s forms, the viscosity of the medium must be effectively zero so that there is negligible transformation of energy between acoustic pressure waves and shear waves.

2.1.2 Frequency Domain Formulations

Because (2.1) is linear in $\Psi(\mathbf{x}, t)$, an equivalent description of the emanations can be formulated in the temporal frequency domain §2.2.6. The emanations can then be described by $\psi(\mathbf{x}, k)$ which is the temporal Fourier transform of $\Psi(\mathbf{x}, t)$, The differential equation for $\psi(\mathbf{x}, k)$ is obtained by Fourier transforming (2.1), thereby producing the Helmholtz wave equation (Morse and Feshbach, 1953, Chapter 11) equation

$$\left(\nabla^2 + k^2\nu^2(\mathbf{x}) + \mu(\mathbf{x}) - jck\beta(\mathbf{x})\right)\psi(\mathbf{x}, k) = 0 \tag{2.2}$$

Although the wave equations (2.1) and (2.2) are exactly equivalent, the latter is in a important sense more versatile because the three constitutive parameters appearing in (2.1) can be combined in the frequency domain into a single generalised constitutive parameter, written here as $\chi = \chi(\mathbf{x}, k)$. Then, (2.2) can be simplified, and simultaneously generalised, to

$$\left(\nabla^2 + k^2\chi^2(\mathbf{x}, k)\right)\psi(\mathbf{x}, k) = 0 \tag{2.3}$$

where

$$\chi^2(\mathbf{x}, k) = \nu^2(\mathbf{x}) - j\beta(\mathbf{x})/k + \mu(\mathbf{x})/k^2 \tag{2.4}$$

The first term on the right hand side of (2.4), which is independent of k, determines the speed of the wave motion at each point in the medium. The square root of this term is what is known historically as the refractive index.

In free space $\mu(\mathbf{x}) = \beta(\mathbf{x}) = 0$ and $\nu(\mathbf{x}) = 1$. The wave equation then becomes

$$\nabla^2\psi(\mathbf{x}, k) + k^2\psi(\mathbf{x}, k) = 0 \tag{2.5}$$

which is known as the free-space wave equation.

2.1.3 Comparison of Time and Frequency Domain Formulations

This section compares time domain and frequency domain formulations thereby allowing the advantages and disadvantages of each formulation to be highlighted.

Time domain formulations are partial hyperbolic differential equations whereas frequency domain formulations are elliptic partial differential equations (Morse and Feshbach, 1953, Chapter 6). Because hyperbolic equations admit solutions which are discontinuous, the notion of causality, as exemplified by the propagation of pulses, can be incorporated into time domain formulations (Garabedian, 1964, Chapter 5). This represents a powerful argument in favour of time domain formulations. However, a disadvantage which often outweighs this is that discontinuous solutions can be highly

unstable numerically. On the other hand, frequency domain formulations can allow several different characteristics of a medium to be combined into a single constitutive parameter, as is manifested by (2.4). For example, dispersion and attenuation, characterised by $\mu(\mathbf{x})$ and $\beta(\mathbf{x})$ respectively, which have to be inserted separately into time domain formulations, can be combined in the frequency domain by permitting $\chi(\mathbf{x}, k)$ to be complex and frequency dependent.

From a theoretical point of view, time domain and frequency domain formulations are equivalent. However, the numerical difficulties arising from time domain approaches can be daunting. Consequently, as is in accordance with the majority of the current literature on inverse problems, this thesis concentrates exclusively on frequency domain formulations.

2.2 Supplementary Definitions and Mathematical Techniques

This section is a collection of concepts, definitions and mathematical techniques which are made use of later in the thesis. As is conventional, "right hand side" and "left hand side" are abbreviated to RHS and LHS respectively, and Re and Im are the abbreviations for "real part" and "imaginary part" respectively.

2.2.1 Partitioning of Wave Motion

The incident and scattered fields are partitioned according to (1.4). The most convenient definition of the incident field is to make it a solution of (2.5) in all of space apart from Υ_{inc}, where the sources of the incident field reside. This then forces $\psi_i(\mathbf{x}, k)$ to be a solution of

$$\left(\nabla^2 + k^2\right) \psi_i(\mathbf{x}, k) = 0 \tag{2.6}$$

everywhere except within Υ_{inc}, so that (2.3) becomes

$$\left(\nabla^2 + k^2\right) \psi_s(\mathbf{x}, k) = -k^2 \left[\chi(\mathbf{x}, k)^2 - 1\right] \psi(\mathbf{x}, k) \tag{2.7}$$

Because LHS (2.7) has the same form as that of the free-space wave equation (2.5), the negative of the RHS can be considered to be a density of equivalent sources embedded in free-space. The field emanating from these sources constitutes the scattered field.

A particularly convenient type of incident radiation is the plane wave. It can be written as

$$\psi_i(\mathbf{x}, k) = e^{-j\mathbf{k}\cdot\mathbf{x}} \tag{2.8}$$

where the direction of propagation is specified by the unit vector $\hat{k} = \mathbf{k}/k$. Surfaces of constant complex amplitude are specified as

$$\mathbf{k} \cdot \mathbf{x} = constant \tag{2.9}$$

Such surfaces of course coincide with surfaces of constant phase. These surfaces are also known as wavefronts. Complementary to wavefronts are rays. A ray is defined as a curve perpendicular to each of the wavefronts which it intersects. Rays represent the paths along which energy propagates.

For the reasons given in §1.5, inverse methods are always illustrated numerically in two dimensions in this thesis. Specific two-dimensional forms of the plane wave are consequently needed. If the arbitrary point $\mathbf{x}$ is accorded the cylindrical coordinates $(\rho;\phi)$ then (2.8) can be expressed a

$$\psi(\rho;\phi,k) = e^{-jk\rho\cos(\phi-\theta)} \tag{2.10}$$

where θ specifies the direction of propagation of the plane wave.

A multipole expansion can also be invoked to represent the planar incident field. When this is done, (2.8) becomes (Jones, 1964, Chapter 8)

$$\psi_i(\mathbf{x},k) = \psi_i(\rho;\phi,k) = \sum_{m=-\infty}^{\infty} a_m J_m(k\rho)e^{jm\phi} \tag{2.11}$$

where $J_m(\cdot)$ is the Bessel function of the first kind of order m, and the a_m depend on the direction of propagation through $a_m = (-j)^m e^{-jm\theta}$.

2.2.2 Far-field (Fraunhofer region) Conditions

In perhaps the great majority of practical applications of inverse scattering, Υ_{obs} lies in what is called the far-field, or Fraunhofer region, of Υ_-. Consider an arbitrary point $P \in \Upsilon_{obs}$ and any two points P' and P'' lying in Υ_-. The essence of the far-field approximation is that the straight lines $P'P$ and $P''P$ can be taken with negligible error to be parallel. This means that the actual value of $\Delta = |P'P - P''P|$ is negligibly different from what it would be if P was infinitely distant from Υ_-. What actually matters, of course, is how large Δ/λ is, where λ is the free-space wavelength of the wave motion. If L is the maximum transverse dimension of Υ_-, as seen from P, then $\Delta \approx L^2/8R$, where $R = PP'$ averaged over all $P' \in \Upsilon_-$. Radio-engineering folk-lore (i.e. what antenna engineers have found by experience, over the past 50 or so years, to be appropriate) is that Δ should not exceed $\lambda/16$, which implies the far-field distance is $2L^2/\lambda$ (Goodman, 1975, Chapter 1; Blake, 1984, Chapter 6) If $\lambda/16$ is considered too large a value for Δ in a certain application then the distance must be increased, but not necessarily by much since Δ varies as $1/R^2$.

2.2.3 Characterisation of Scattering

The relative strength of the scattered field compared to the incident field is an important factor in many situations. It is convenient to classify the "strength" of the scattering as very weak, weak or strong. To formalise the classification it is necessary to define the quantities

$$A_{diff}(\mathbf{x},k) = |\psi(\mathbf{x},k)| - |\psi_i(\mathbf{x},k)| \tag{2.12}$$

which is here called the amplitude increment, and

$$\mathrm{Ph}_{diff}(\mathbf{x},k) = \angle\psi(\mathbf{x},k) - \angle\psi_i(\mathbf{x},k) \tag{2.13}$$

which is called the phase increment, where $\angle q$ denotes the phase of q. The incremental phase change $\Delta(\mathbf{x},\mathbf{x}_1,k)$ between points $\mathbf{x}$ and $\mathbf{x}_1$ is defined to be

$$\Delta(\mathbf{x},\mathbf{x}_1,k) = \mathrm{Ph}_{diff}(\mathbf{x},k) - \mathrm{Ph}_{diff}(\mathbf{x}_1,k) \tag{2.14}$$

Very weak scattering can then be said to occur when $A_{diff} \approx 0$ throughout Υ_- and $\Delta(\mathbf{x}, \mathbf{x}_1, k) < \frac{\pi}{2}$ for all pairs of points $\mathbf{x}$ and $\mathbf{x}_1$ at which rays of ψ_i enter and leave Υ_-. Recall that the concept of rays is introduced descriptively in §1.5. It is made more precise in §2.2.1.

Weak scattering is defined similarly to very weak scattering except that $\mathbf{x}$ and $\mathbf{x}_1$ are points on rays of ψ_i separated by λ.

Strong scattering occurs when, within an appreciable fraction of Υ_-, A_{diff} differs significantly from zero and/or $\Delta(\mathbf{x}, \mathbf{x}_1, k)$ is appreciably greater than $\frac{\pi}{2}$ for points $\mathbf{x}$ and $\mathbf{x}_1$ on rays of ψ_i separated by λ.

2.2.4 The Green's Function

The Green's function is defined as the particular solution of the equation

$$(\nabla^2 + k^2\nu^2) g(\mathbf{x}, \mathbf{x}_1, k\nu) = -\delta(\mathbf{x} - \mathbf{x}_1) \tag{2.15}$$

where the quantity $g(\mathbf{x}, \mathbf{x}_1, k\nu)$ is called the Green's function for the operator $(\nabla^2 + k^2\nu^2)$ (Morse and Feshbach, 1953, Chapter 6). Physically, it represents the field at the point $\mathbf{x}$ radiated from a point source of unit strength at $\mathbf{x}_1$. Writing $|\mathbf{x} - \mathbf{x}_1| = r$, it transpires that the two- and three-dimensional forms of $g(\mathbf{x}, \mathbf{x}_1, k\nu)$ are (Jones, 1964, Chapter 1)

$$g(\mathbf{x}, \mathbf{x}_1, k\nu) = \begin{cases} \frac{-j}{4} H_0^{(2)}(k\nu r) & \mathbf{x} \in \mathbf{R}^2 \\ \\ \frac{e^{-jk\nu r}}{4\pi r} & \mathbf{x} \in \mathbf{R}^3 \end{cases} \tag{2.16}$$

where $H_0^{(2)}(\cdot)$ is the cylindrical Hankel function of the second kind of order zero, and $\mathbf{R}^K$ denotes the set of points in K-dimensional space.

For the reasons given in §1.5, it can be appropriate as well as convenient to illustrate numerical aspects of inverse methods in two dimensions. If the arbitrary points $\mathbf{x}$ and $\mathbf{x}_1$ are accorded the cylindrical polar coordinates $(\rho; \phi)$ and $(r; \theta)$ in two dimensions then a particularly useful multipole expansion of $g(\mathbf{x}, \mathbf{x}_1, k\nu)$, for $\mathbf{x}, \mathbf{x}_1 \in \mathbf{R}^2$, is (Watson, 1966, Chapter 11)

$$g(\rho; \phi, r; \theta, k\nu) = \frac{-j}{4} \sum_{m=-\infty}^{\infty} H_m^{(2)}(k\nu\rho) J_m(k\nu r) e^{jm(\phi - \theta)} \tag{2.17}$$

with $\rho \geq r$, where $J_m(\cdot)$ and $H_m^{(2)}(\cdot)$ denote the Bessel function of the first kind and the Hankel function of the second kind, respectively, of order m. Note that the Bessel and Hankel functions must be interchanged when $r < \rho$.

The form of the Green's function simplifies in the far-field of Υ_-, wherein (by definition) $H_m^{(2)}(k\nu\rho)$ can be replaced by its asymptotic expansion (Abramowitz and Stegun, 1965, Chapter 9)

$$H_m^{(2)}(k\nu\rho) = \sqrt{\frac{2}{\pi k\nu\rho}} e^{-j(k\nu\rho - \frac{1}{2}m\pi - \frac{1}{4}\pi)} \tag{2.18}$$

Note that this can be regarded as an alternative definition of the far-field. Consequently,

the Green's function can be written in the far-field as

$$g(\mathbf{x},\mathbf{x}_1,k\nu)=\begin{cases} e^{-j\frac{3\pi}{4}}\frac{e^{-jk\nu(|\mathbf{x}|-\mathbf{x}_1\cdot\hat{\mathbf{x}})}}{\sqrt{8\pi k\nu|\mathbf{x}|}} & \mathbf{x}\in\mathbf{R}^2 \\ \\ \frac{e^{-jk\nu(|\mathbf{x}|-\mathbf{x}_1\cdot\hat{\mathbf{x}})}}{4\pi|\mathbf{x}|} & \mathbf{x}\in\mathbf{R}^3 \end{cases} \tag{2.19}$$

where the unit vector $\hat{\mathbf{x}}$ is defined by $\hat{\mathbf{x}}=\mathbf{x}/|x|$. The formulas (2.18) and (2.19) emphasise that, in the far-field of Υ_-, the curvature of the spherical wavefronts becomes negligible and $g(\mathbf{x},\mathbf{x}_1,k)$ behaves like a plane wave. The terms in the denominators of the expressions on RHS (2.19) ensure that power is conserved as the wave propagates through space.

Because of (2.19) the far-field form of $\psi_s(\mathbf{x},k)$ is usually expressed as $\psi_{ffs}(\hat{\mathbf{x}},k)$, which is defined by

$$\psi_s(\mathbf{x},k)=\begin{cases} \frac{e^{-jk|\mathbf{x}|}}{\sqrt{8\pi k|\mathbf{x}|}}\psi_{ffs}(\hat{\mathbf{x}},k) & \mathbf{x}\in\mathbf{R}^2 \\ \\ \frac{e^{-jk|\mathbf{x}|}}{4\pi|\mathbf{x}|}\psi_{ffs}(\hat{\mathbf{x}},k) & \mathbf{x}\in\mathbf{R}^3 \end{cases} \tag{2.20}$$

where $\hat{\mathbf{x}}$ is the unit vector in the direction of $\mathbf{x}$. $\psi_{ffs}(\hat{\mathbf{x}},k)$ is that part of $\psi(\mathbf{x},k)$ which only depends on $\hat{\mathbf{x}}$ and has no dependence on $|\mathbf{x}|$.

2.2.5 Boundary, Radiation, Initial, Jump and Finiteness Conditions

Besides the differential equation (2.7), certain supplementary equations, which are dependent on the situation being considered, must also be invoked when attempting to solve the direct and inverse problems. Four of these supplementary equations are called boundary conditions, initial conditions, radiation conditions and jump conditions. They are treated in texts describing wave phenomena (Morse and Feshbach, 1953, Chapter 6; Jones, 1964, Chapter 1) and are stated whenever they are needed in later Chapters of this thesis and in subsequent sections of this Chapter. There is a fifth supplementary equation, called the finiteness condition, which states that the energy of the emanations in any finite region of Υ must be finite (Jones, 1964, Chapter 9).

2.2.6 The Fourier Transform

Because it makes its presence felt in several subsequent Chapters, the Fourier transform and its inverse are now defined. Given a K-dimensional function $f(\mathbf{x})$, its K-dimensional spectrum $F(\mathbf{u})$ is defined by (Bracewell, 1978, Chapter 2)

$$F(\mathbf{u})=\int\!(K)\!\int_{-\infty}^{\infty} f(\mathbf{x})\,e^{-j\mathbf{u}\cdot\mathbf{x}}\,dV(\mathbf{x}) \tag{2.21}$$

where $\mathbf{u}$ is a K-dimensional position vector reciprocal to $\mathbf{x}$, the notation $\int\!(K)\!\int$ represents a K-dimensional integral and $dV(\mathbf{x})$ is the volume element in the K-dimensional space. The inverse formula to (2.21) is

$$f(\mathbf{x})=\frac{1}{(2\pi)^K}\int\!(K)\!\int_{-\infty}^{\infty} F(\mathbf{u})\,e^{j\mathbf{u}\cdot\mathbf{x}}\,dV(\mathbf{u}) \tag{2.22}$$

The functions $f(\mathbf{x})$ and $F(\mathbf{u})$ constitute a Fourier transform pair, a convenient explicit shorthand notation for which is

$$F(\mathbf{u}) = \mathcal{F}_K[f(\mathbf{x})] \tag{2.23}$$

and the inverse

$$f(\mathbf{x}) = \mathcal{F}_K^{-1}[F(\mathbf{u})] \tag{2.24}$$

It is appropriate to regard $\mathbf{x}$ and $\mathbf{u}$ as position vectors of arbitrary points in image space and Fourier space, respectively.

A special one-dimensional form of the Fourier transform is also worthwhile defining. It is known as the Temporal Fourier transform and is written as (Bracewell, 1978, Chapter 2)

$$\psi(\mathbf{x}, k) = \int_{-\infty}^{\infty} \Psi(\mathbf{x}, t) e^{-jkt} dt \tag{2.25}$$

which can be inverted to give

$$\Psi(\mathbf{x}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(\mathbf{x}, k) e^{jkt} dk \tag{2.26}$$

$\psi(\mathbf{x}, k)$ is referred to throughout this thesis as the temporal frequency spectrum of $\Psi(\mathbf{x}, t)$ (Goodman, 1968). The wavelength λ corresponding to the wave number k is $\frac{2\pi}{k}$.

2.2.7 Sampling and the Sampling Theorem

Before data can be operated on in a digital computer, they must first be sampled. The sample of a function $f(\mathbf{x})$ at the three-dimensional point $\mathbf{x}_{m,n,p}$ is here defined to be

$$f_{m,n,p} = \int(3) \int_{-\infty}^{\infty} f(\mathbf{x}) \delta(\mathbf{x}_{m,n,p} - \mathbf{x}) dV(\mathbf{x}) \tag{2.27}$$

Of particular interest are the samples of functions which are bandlimited. Consider the bandlimited function $f(\mathbf{x})$ whose Fourier transform $F(\mathbf{u})$ satisfies

$$F(\mathbf{u}) = 0 \quad \text{except for} \quad \begin{cases} -2\pi L_1 < u_1 < 2\pi L_1 \\ -2\pi L_2 < u_2 < 2\pi L_2 \\ -2\pi L_3 < u_3 < 2\pi L_3 \end{cases} \tag{2.28}$$

where u_1, u_2 and u_3 are Cartesian components of $\mathbf{u}$. If

$$\mathbf{x}_{m,n,p} = \left(\frac{m}{L_1}, \frac{n}{L_2}, \frac{p}{L_3} \right) \tag{2.29}$$

then, by the sampling theorem (Bates and McDonnell, 1989, Chapter 2), $f(\mathbf{x})$ can be exactly reconstructed from its samples by invoking the formula

$$f(\mathbf{x}) = \sum_{m,n,p=-\infty}^{\infty} f_{m,n,p} \text{sinc}(L_1 x_1 - m) \text{sinc}(L_2 x_2 - n) \text{sinc}(L_3 x_3 - p) \tag{2.30}$$

where

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{2.31}$$

2.2.8 Fredholm Integral Equations

If $\mathbf{x}$ and $\mathbf{y}$ are position vectors of arbitrary points in regions $\mathcal{A}$ and $\mathcal{B}$, respectively, of K-dimensional space, and if an unknown function $u(\mathbf{x})$ is related to a given function $h(\mathbf{x})$ and a known kernel $k(\mathbf{x},\mathbf{y})$ through

$$h(\mathbf{x}) = C\,u(\mathbf{x}) + \int(K)\int_{\mathcal{B}} u(\mathbf{y})\,k(\mathbf{x},\mathbf{y})\,dV(\mathbf{y}) \tag{2.32}$$

where C is a constant then (2.32) is known as a Fredholm integral equation of the second kind (Morse and Feshbach, 1953, Chapter 8). It becomes an equation of the first kind when $C = 0$.

To solve (2.32) it is necessary to transform it (approximately) into a finite number of linear algebraic equations (Bates *et al.*, 1990). To this end, sets $\{\Phi_m(\mathbf{x}); m = 1,2,\cdots,M\}$ and $\{\Psi_n(\mathbf{x}); n = 1,2,\cdots,N\}$ of basis functions are introduced, with $g(\mathbf{x})$ and $u(\mathbf{x})$ being respectively defined by

$$h(\mathbf{x}) = \sum_{m=1}^{M} A_m \Phi_m(\mathbf{x}) \quad \text{and} \quad u(\mathbf{x}) = \sum_{n=1}^{N} B_n \Psi_n(\mathbf{x}) \tag{2.33}$$

where the A_m and B_n are constants called expansion coefficients. Since $g(\mathbf{x})$ is given, the A_m can immediately be evaluated, provided the $\Phi_m(x)$ are independent throughout $\mathcal{A}$. The calculation is eased if the $\Phi_m(x)$ are biorthogonal (Morse and Feshbach, 1953, Chapter 8) throughout $\mathcal{A}$ to another set $\{\Theta_m(x); m = 1,2,\cdots,M\}$ of basis functions, i.e.

$$\int(K)\int_{\mathcal{A}} \Phi_m(x)\,\Theta_n(x)\,w(x)\,dV(x) = C_m \delta_{m,n} \tag{2.34}$$

where $w(\mathbf{x})$ is an appropriate weight function, the C_m are constants and δ_{mn} is the kronecker delta which is defined for integers m and n by

$$\delta_{mn} = 1 \quad \text{for} \quad m = n \quad \text{and} \quad \delta_{mn} = 0 \quad \text{for} \quad m \neq n. \tag{2.35}$$

On defining

$$k_{m,n} = \int(K)\int_{\mathcal{A}}\int(K)\int_{\mathcal{B}} \Psi_n(\mathbf{y})\,\Theta_m(\mathbf{x})\,k(\mathbf{x},\mathbf{y})\,w(\mathbf{x})\,dV(x)\,dV(y) \tag{2.36}$$

and

$$u_{m,n} = \int(K)\int_{\mathcal{A}} \Psi_n(x)\,\Theta_m(x)\,w(x)\,dV(x) \tag{2.37}$$

it can be seen that multiplying (2.32) through by $(\Theta_m(\mathbf{x})w(\mathbf{x}))$ and integrating over $\mathcal{A}$ gives

$$C_m A_m = \sum_{n=1}^{N} B_n\,[C\,u_{m,n} + k_{m,n}] \quad \text{for} \quad m = 1,2,\cdots,M \tag{2.38}$$

which is conveniently rewritten as

$$\mathbf{A} = [k]\,\mathbf{B} \tag{2.39}$$

where $\mathbf{A}$ and $\mathbf{B}$ are vectors, whose components are the A_m and B_n respectively, and $[k]$ is the $M \times N$ matrix whose mn^{th} element is $(C\,u_{mn} + k_{mn})$.

Since the B_n are the only unknowns in (2.39), they represent the solution to (2.32). When $M > N$, premultiplying (2.39) by the transpose $[k]^T$ of $[k]$ and then by the inverse of $[k]^T[k]$, gives

$$\mathbf{B} = \left([k]^T[k]\right)^{-1}[k]^T\mathbf{A} \tag{2.40}$$

which, provided $[k]^{-1}$ exists, reduces when $M = N$ to

$$\mathbf{B} = [k]^{-1}\mathbf{A} \tag{2.41}$$

Note that (2.40) is equivalent to the least squares solution for **B** (Atkinson, 1978, Chapter 7). Remember also that there is no solution when the system of equations (2.38) is underdetermined, by which is meant that $M < N$. While standard software can be invoked to implement either (2.40) or (2.41), it is important to keep in mind that the only practical means of assessing the stability of any such solution is by increasing M and N and inspecting the degree of numerical convergence resulting therefrom (Jones, 1979, Chapter 5). Analysis can sometimes (but by no means always) establish whether convergence is to be expected, but the all important practical consideration of the rate of convergence can only be found by computational experimentation (Mittra, 1973, Chapter 2).

The matrix inversions implied in (2.40) and (2.41) tend to be less sensitive to errors in the data (i.e erroneous values of the A_m) when $C \neq 0$ (i.e for Fredholm equations of the second, as opposed to the first kind) (Arfken, 1985, Chapter 16). This sensitivity of equations of the first kind is much less significant for situations in which both u_{mn} and k_{mn} vanish for $m \neq n$. Matrix inversion is then unnecessary because each B_n is directly proportional to A_n with the constant of proportionality depending in general upon n. Integral transforms represent important examples of such Fredholm equations of the first kind. Note that the Fourier integral (2.21) is equivalent to (2.32) with $C = 0$, when $F(\cdot)$, $f(\cdot)$ and $\exp(\cdot)$ are identified with $g(\cdot)$, $u(\cdot)$ and $k(\cdot)$ respectively. (2.22) emphasis that (2.21) can be immediately inverted without recourse to matrix inversion. It is finally worth emphasising that discrete Fourier transforms (DFT) (Bracewell, 1978) are as immediately invertible as are continuous Fourier transforms, with spectacular numerical efficiency due to the FFT (fast Fourier transform) algorithm (Cooley and Tukey, 1965; Brigham, 1974).

2.2.9 Distance Measures

It is often useful to specify a generalised distance between two vectors, **A** and **B** say. Three distance measures which are often invoked are the one, two and infinity norms (Atkinson, 1978), which are defined below. The notation A_m is adopted for the m^{th} component of the vector **A**.

The one norm is defined as the sum of the absolute values of the differences between the components of **A** and **B**:

$$||\mathbf{A} - \mathbf{B}||_1 = \sum_m |A_m - B_m| \tag{2.42}$$

The two norm is defined as the conventional Euclidean distance between **A** and **B**:

$$||\mathbf{A} - \mathbf{B}||_2 = \left[\sum_m |A_m - B_m|^2\right]^{\frac{1}{2}} \tag{2.43}$$

The infinity norm is the absolute value of the maximum difference between components of **A** and **B**:

$$||\mathbf{A} - \mathbf{B}||_\infty = \max_m |A_m - B_m| \tag{2.44}$$

Distance measures can be invoked for the quantitative specification of the error between a quantity **A** and an estimate **B** of it. The three distance measures defined above are conveniently combined into the composite norm error

$$\epsilon_{comp} = ||\mathbf{A} - \mathbf{B}||_n \tag{2.45}$$

where $n = 1,2$ or ∞. It is convenient to introduce the terminology: relative norm n error, which is defined by

$$\epsilon = \frac{||\mathbf{A} - \mathbf{B}||_n}{||\mathbf{A}||_n} \tag{2.46}$$

where $n = 1,2$ or ∞.

2.2.10 Types of Scattering Data

Four different types of scattering data can be acquired in Υ_{obs}. They are here called complex amplitude, magnitude, bistatic cross-section and total cross-section. These types of scattering data, which are defined below, are each identified by attaching a subscript (c, m, b and t respectively) to the symbol D.

Scattering data D_c of the type called complex amplitude constitute the whole of $\psi(\mathbf{x}, k)$ for $\mathbf{x} \in \Upsilon_{obs}$. These data comprise both the magnitude and phase of the field:

$$D_c = \psi(\mathbf{x}, k) \quad \text{for} \quad \mathbf{x} \in \Upsilon_{obs} \tag{2.47}$$

Scattering data D_m of the type called magnitude constitute only the magnitude of $\psi(\mathbf{x}, k)$:

$$D_m = |\psi(\mathbf{x}, k)| \quad \text{for} \quad \mathbf{x} \in \Upsilon_{obs} \tag{2.48}$$

Bistatic and total cross-section scattering data can only be acquired when the incident field is planar and when Υ_{obs} is in the far-field of Υ_-. Scattering data D_b of the type called bistatic cross-section constitute

$$D_b = \varsigma(\hat{\mathbf{x}}) = |\psi_{ffs}(\hat{\mathbf{x}}, k)|^2 \tag{2.49}$$

where $\varsigma(\mathbf{x})$ is known as the bistatic cross-section (Bowman *et al.*, 1969, Chapter 1). Scattering data D_t of the type called total cross-section constitute

$$D_t = \varsigma_T = \int_\Gamma \varsigma(\hat{\mathbf{x}}) d\Gamma(\hat{\mathbf{x}}) \tag{2.50}$$

where Γ is the surface of unit radius and ς_T is called the total cross-section (Bowman *et al.*, 1969, Chapter 1). The bistatic and total cross-sections are related through

$$\varsigma_T = \text{Im}\left(\varsigma(\hat{\mathbf{k}})\right) \tag{2.51}$$

which is also known as the forward scattering theorem (Bowman *et al.*, 1969, Chapter 1).

2.2.11 Noise Levels for Numerical Examples

When illustrating direct and inverse scattering techniques with numerical examples, it is appropriate to add realism by contaminating the scattering data with noise. Before the noise level can be quantitatively stated, the probability density distribution (pdf) of the noise must be specified. In this thesis, only rectangular and Gausssian pdfs are considered. For each pdf, the noise level is defined by the ratio

$$r = \frac{3s}{||D||_\infty} \tag{2.52}$$

where s is the standard deviation of the pdf and $||D||_\infty$ is the infinity norm (see §2.2.9) of the scattering data.

It can be convenient to state noise levels in percentages. So, if $\alpha = 100r$, the noise level is said to be $\alpha\%$, either rectangularly or Gaussianly distributed.

Chapter 3

Review of Solutions to Direct Scattering Problems

This Chapter is concerned with the direct problem, which elucidates how emanations $\psi(\mathbf{x}, k)$ interact with a body of known generalised constitutive parameter $\chi(\mathbf{x})$. The solution of the direct problem is an important prerequisite for solving the inverse problem (see §1.4). Direct problems which involve wave-like emanations are analysed in this Chapter. These emanations are the same as those featuring in the inverse scattering problems considered in Chapters 4 and 5.

The direct problem is formally defined in §1.3.2 as the mapping

$$\psi_s(\mathbf{x}, k) = \Lambda\left(\chi(\mathbf{x}), \psi(\mathbf{x}, k)\right) \tag{3.1}$$

where Λ, which models the physical processes involved, depends on the type of emanations under consideration and on the way they interact with the constitutive parameter (see §2.1). A general explicit expression for $\psi_s(\mathbf{x}, k)$ in terms of $\chi(\mathbf{x})$ cannot however usually be stated.

When the direct problem is considered in isolation, numerical calculations of $\psi_s(\mathbf{x}, k)$ for a particular $\chi(\mathbf{x})$ and for given emanations $\psi_i(\mathbf{x}, k)$ can be effected by resorting to computational solutions of either a differential or an integral equation. A complete discussion of all known techniques would require many volumes of text in their own right. In this thesis, therefore, the emphasis is on solutions to direct problems which are computationally efficient. They can then then be incorporated into solutions to inverse problems of interest in practical situations.

A common approach to making direct problems computationally tractable is to employ various approximations, which sometimes can lead to expressions that allow $\chi(\mathbf{x})$ to be analytically inverted. The study of such approximate solutions can enhance one's conceptual understanding of the physics of wave propagation. This can help one to appreciate why some approximate solutions to inverse problems fail and how they might be usefully modified. An understanding of the physics of wave propagation may also lead to the development of new inverse methods, which fall between the exact, computationally slow and the approximate, computationally fast solutions.

It is sometimes impossible to find any viable approximate approach to solving the direct problem. This may be because appropriate approximations are too elusive or those approximations which do suggest themselves are so imprecise as to be useless. In these cases it is necessary to resort to numerical methods which solve direct problems

to arbitrary accuracy. It is, however, important to attempt to minimise the total numerical effort required to solve any particular direct problem.

This Chapter reviews computationally efficient solutions to direct problems. Approximate solutions of the direct problem which are valid at high frequencies are considered in §3.1. In particular, the high frequency solutions known as ray tracing, geometrical optics, the geometrical theory of diffraction and the WKB methods are explained. Solution of the direct problem using the volume source formulation is introduced in §3.2. Approximations to the volume source formulation which lead to the Rayleigh-Gans (Born) and Rytov approximations are detailed in §3.2.1 and §3.2.2. Full wave solutions to the direct problem are discussed in §3.3 through §3.5. Finally, in §3.6 a recently developed technique for solving the direct problem known as the on-surface radiation condition is discussed. Numerical examples are presented throughout this Chapter to illustrate the various methods treated herein.

All theoretical developments are presented in three dimensions but, for computational economy, all of the illustrative examples presented in this Chapter are two-dimensional.

3.1 High Frequency Approximations

In many situations of practical interest, the refractive index of the medium only varies appreciably over distances which are large compared to the wavelength of the emanations. In fact, this is true for all media if the frequency of the emanations is high enough. Consequently, approximations which take advantage of this are known as high frequency approximations (Jones, 1964, Chapter 6). In such situations it can be assumed that the field behaves locally as a plane wave but varies globally over the medium.

A plane wave can be written as

$$\psi_i(\mathbf{x}, k) = e^{-j\mathbf{k}\cdot\mathbf{x}} \tag{3.2}$$

where the direction of propagation is specified by the unit vector $\hat{k} = \mathbf{k}/k$. Surfaces of constant amplitude are specified as

$$\mathbf{k}\cdot\mathbf{x} = constant \tag{3.3}$$

Such surfaces of course coincide with surfaces of constant phase. These surfaces are also known as wavefronts.

Whenever high frequency approximations are valid, it can be expected that the wavefronts are retarded differently in different parts of the medium, implying that the field should be expressed in some such form as

$$\psi = A(\mathbf{x})e^{-jkS(\mathbf{x})} \tag{3.4}$$

where $A(\mathbf{x})$ and $S(\mathbf{x})$ are real, so that $kS(\mathbf{x})$ is the phase of the field and $A(\mathbf{x})$ is its amplitude. If (3.4) is substituted into the homogeneous Helmholtz equation (2.3) then

$$\begin{aligned} 0 \;=\; & k^2(\nu^2 - \nabla S\cdot\nabla S)A \\ & -jk(A\nabla^2 S + 2\nabla A\cdot\nabla S) \\ & +\nabla^2 A \end{aligned} \tag{3.5}$$

When any high frequency approximation is valid, k is large by definition, and significant variations in $A(\mathbf{x})$ only occur over many wavelengths. Consequently, the most significant term in (3.5) is the coefficient of k^2. Equating this coefficient to zero gives

$$\nabla S \cdot \nabla S = \nu^2(\mathbf{x}) \tag{3.6}$$

which is known as the eikonal equation (Jones, 1964). Along with the initial conditions, (3.6) defines $S(\mathbf{x})$ in terms of $\nu(\mathbf{x})$, the refractive index of the medium.

$A(\mathbf{x})$ can be estimated by equating the coefficient of k in (3.5) to zero:

$$A\nabla^2 S + 2\nabla A \cdot \nabla S = 0 \tag{3.7}$$

which is known as the first transport equation. When $\nu(\mathbf{x})$ varies predominantly along a single Cartesian direction, the solution of (3.7) is

$$A(\mathbf{x}) = C\nu^{-\frac{1}{2}}(\mathbf{x}) \tag{3.8}$$

with C being an arbitrary constant.

In general, (3.7) cannot be solved explicitly, so that approximate expressions for $A(\mathbf{x})$ must be sought.

Methods such as ray tracing, geometrical optics, the geometrical theory of diffraction and the WKB technique can be variously invoked to solve (3.6) and (3.7) for $S(\mathbf{x})$ and $A(\mathbf{x})$ respectively.

3.1.1 Ray Tracing

Ray tracing is here defined to be any procedure for calculating $S(\mathbf{x})$ and $A(\mathbf{x})$, from (3.6) and (3.7) in a medium whose refractive index is continuous.

A geometrical ray is defined as a curve, everywhere perpendicular to the wavefronts, representing the direction in which energy propagates. Because any surface $S(\mathbf{x}) =$ constant represents a wavefront, $\nabla S(\mathbf{x})$ must be normal to the wavefront. Thus, $\nabla S(\mathbf{x})$ determines the direction of each ray. Inspection of (3.6) reveals that the magnitude of $\nabla S(\mathbf{x})$ is $\nu(\mathbf{x})$. Therefore, the differential equation for a ray is

$$\frac{d\mathbf{r}(s)}{ds} = \frac{\nabla S(\mathbf{x})}{\nu(\mathbf{x})} \tag{3.9}$$

where $\mathbf{r}(s)$ is the position vector of an arbitrary point on a ray and s is arc length along the ray.

An alternative derivation of the ray path through a medium can be based on Fermat's principle, which states that energy propagates along the path for which the time of flight is a minimum (Born and Wolf, 1970, Chapter 3). The normalised time of flight along a path between the points P_1 and P_2 is

$$I = \int_{P_1}^{P_2} \nu(\mathbf{x})\, ds \tag{3.10}$$

where I is also known as the acoustical or electrical path length. The actual ray path is that which minimises the integral in (3.10). The Euler equations (Arfken, 1985,

Chapter 17), which constitute a necessary condition for the existence of a minimum value for (3.10), yield the differential equation

$$\nu(\mathbf{x})\frac{d^2\mathbf{r}(s)}{ds^2} + \left(\nabla\nu(\mathbf{x})\frac{d\mathbf{r}(s)}{ds}\right)\frac{d\mathbf{r}(s)}{ds} = \nabla\nu(\mathbf{x}) \tag{3.11}$$

where $\mathbf{r}(s)$ is the position vector of an arbitrary point on the ray distant s from P_1, along the ray. Andersen and Kak (1982) show that (3.11) can be manipulated into the form of (3.6), confirming that Fermat's principle is a high frequency approximation to the Helmholtz wave equation.

Algorithms for calculating the ray paths in a continuous medium are based on any one of the four equations (3.6), (3.9), (3.10) or (3.11). Examples of the many ray tracing algorithms which have been developed are those due to Andersen and Kak (1982), Smith *et al.* (1980), Fawcett and Keller (1985), Norton (1987) and Walther (1988).

Once the rays have been traced through the medium, it is possible to calculate the amplitudes of the field propagating along the rays. This is accomplished by invoking conservation of energy flow. Consider the elemental tube of rays shown in fig. 3.1. Because the energy propagates along the rays, no energy is lost across the sides of the tube, confirming that energy flow is conserved along the tube. Consequently

$$|A_1|^2 d\Sigma_1 = |A_2|^2 d\Sigma_2 \tag{3.12}$$

where A_1 and A_2 are the amplitudes of the wavefront at $d\Sigma_1$ and $d\Sigma_2$, respectively, and $d\Sigma_1$ and $d\Sigma_2$ are the elemental areas of the tube at two positions along the tube. So, the variation in amplitude of the field along the tube can be calculated if the amplitude is known at any one position. Generally the radiation pattern of the source of the rays is known, so its amplitude can be used as a reference amplitude.

3.1.1.1 New Ray Tracing Algorithm

In this section, I present a new implementation of the second order algorithm developed by Andersen and Kak (1982). This algorithm is based on expanding $\mathbf{r}(s)$ in a second order Taylor series as

$$\mathbf{r}(s+\Delta s) = \mathbf{r}(s) + \frac{d\mathbf{r}(s)}{ds}\Delta s + \frac{1}{2}\frac{d^2\mathbf{r}(s)}{ds^2}\Delta s^2 \tag{3.13}$$

where Δs is an increment of arc length along the ray from the point represented by $\mathbf{r}(s)$. Eliminating the term $\frac{d^2\mathbf{r}(s)}{ds^2}$ from (3.13) by invoking (3.11) gives

$$\mathbf{r}(s+\Delta s) = \mathbf{r}(s) + \frac{d\mathbf{r}(s)}{ds}\Delta s + \frac{1}{2\nu(\mathbf{x})}\left[\nabla\nu(\mathbf{x}) - \left(\nabla\nu(\mathbf{x})\frac{d\mathbf{r}(s)}{ds}\right)\frac{d\mathbf{r}(s)}{ds}\right]\Delta s^2 \tag{3.14}$$

This formula enables a ray to be traced recursively at discrete points, each separated along the ray by arcs of length Δs, given a starting point $\mathbf{r}(s_0)$ and an initial direction $\frac{d\mathbf{r}(s_0)}{ds}$ for the ray. Note that $\nu(\mathbf{x})$, $\nabla\nu(\mathbf{x})$ and $\frac{d\mathbf{r}(s)}{ds}$, also need to be known at each point $\mathbf{r}(s)$. Since $\nu(\mathbf{x})$ is given in the direct problem, $\nabla\nu(\mathbf{x})$ can be calculated everywhere in the medium. On the other hand, there is no explicit formula for $\frac{d\mathbf{r}(s)}{ds}$, which must be evaluated with the aid of some appropriate numerical procedure for calculating derivatives. Since Andersen and Kak (1982) offer no suggestion as to which procedure

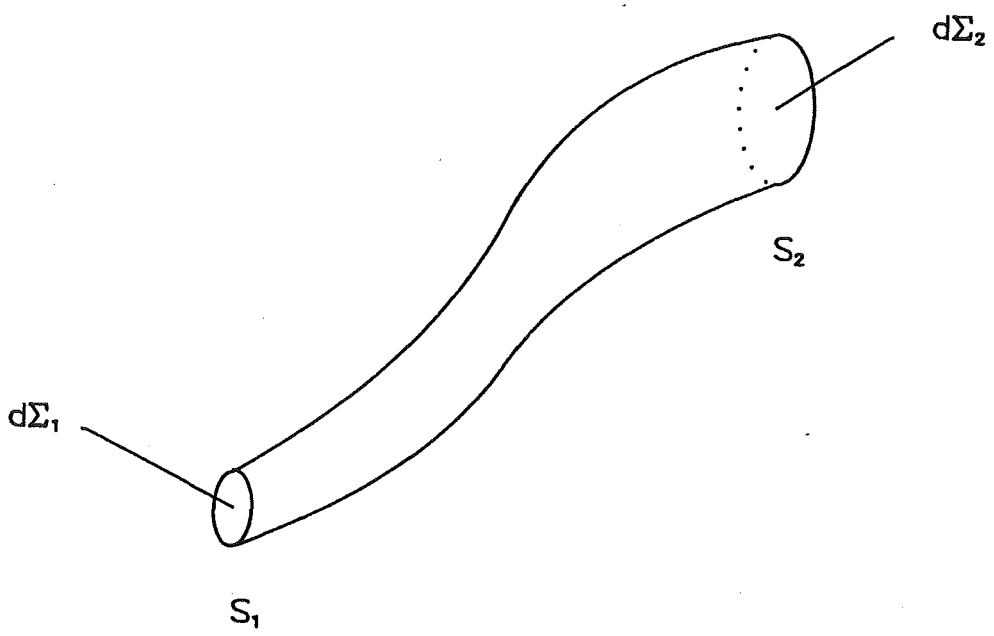


Figure 3.1: An elemental tube of rays of area $d\Sigma_1$ and $d\Sigma_2$ at the wavefront positions S_1 and S_2 respectively.

might be most effective, it seems appropriate to describe (below) an algorithm which I developed specifically for calculating $\frac{d\mathbf{r}(s)}{ds}$.

The algorithm, which is iterative, comprises the following steps:

Step 1: A first estimate of $\frac{d\mathbf{r}(s)}{ds}$ is obtained as

$$\frac{d\mathbf{r}(s)}{ds} = \frac{\mathbf{r}(s-\Delta s) - \mathbf{r}(s)}{\Delta s} \tag{3.15}$$

which is then used to calculate $\mathbf{r}(s+\Delta s)$ using (3.14).

Step 2: From the above value of $\mathbf{r}(s+\Delta s)$ a mid-point estimate of $\frac{d\mathbf{r}(s)}{ds}$ is calculated:

$$\frac{d\mathbf{r}}{ds} = \frac{\mathbf{r}(s-\Delta s) - \mathbf{r}(s+\Delta s)}{2\Delta s} \tag{3.16}$$

Step 3: This new estimate of $\frac{d\mathbf{r}(s)}{ds}$ is then used to calculate a new $\mathbf{r}(s+\Delta s)$

Step 4: Steps 2-3 are repeated until the difference between successive estimates of $\frac{d\mathbf{r}(s)}{ds}$ is less than a pre-set value.

The accuracy of this ray tracing algorithm is conveniently assessed by applying it to "Maxwell's fish-eye" (Barnes and Solomon, 1973), which is the circularly symmetrical refractive index distribution

$$\nu(\mathbf{x}) = \nu_o(1 + \frac{|\mathbf{x}|}{a})^2 \tag{3.17}$$

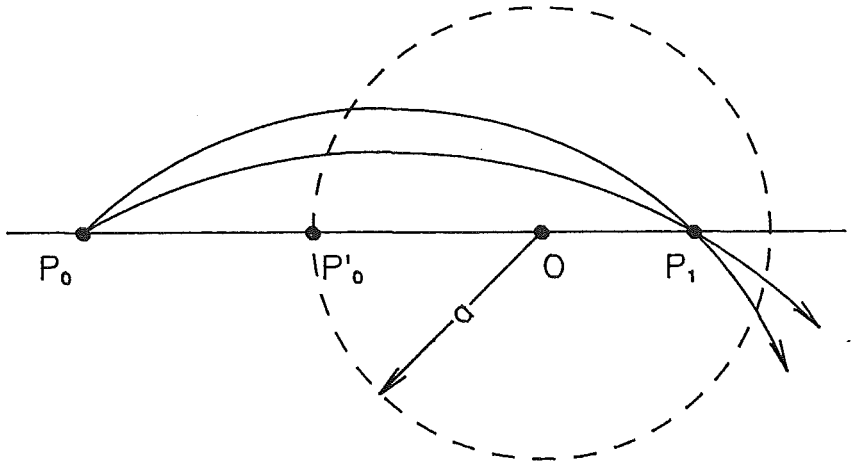


Figure 3.2: Diagram showing rays in Maxwell's "fish-eye". All rays from an arbitrary point P_0 meet at an image point P_1. If a ray is released from P_0', at a distance a from O in a direction perpendicular to the line P_0O, then the ray traces a circle of radius a centred at O.

where $\mathbf{x}$ is the position vector of an arbitrary point, and ν_0 and a are constants. A characteristic property of this distribution is shown in fig. 3.2. All rays from an arbitrary point P_0 meet at an image point P_1 such that the line P_0P_1 intersects O. The distances $|P_0O|$ and $|OP_1|$ are related through $|P_0O| \cdot |OP_1| = a^2$. If a ray is released from P_0', at a distance a from O in a direction perpendicular to the line P_0O, then the ray traces a circle of radius a centred at O. As Andersen and Kak (1982) imply, a severe test for a ray tracing algorithm is to require it to trace such a circle. It is also desirable for the algorithm to do this as efficiently as possible, which is equivalent to employing a minimum number of recursions. The algorithm outlined above requires 250 recursions to close, to within 1 %, a circle having a diameter of 40 Δs. This performance appears superior to that of other published algorithms (Smith *et al.*, 1980; Fawcett and Keller, 1985).

The ray tracing algorithm introduced above can be combined with (3.12) to produce a method of solving the direct problem in the high frequency approximation. In order to compare the accuracy of this method to the Rayleigh-Gans (Born) approximation (§3.2.1) and the eigenfunction expansion method (§3.5) it is necessary to chose some standard object, from which scattering can be conveniently calculated. The chosen object, which is shown in fig. 3.3, is a circular cylinder of radius r_0 of constant refractive index ν_-. Thus, $\nu = \nu_-$ throughout Υ_-, with $\nu = 1$ throughout Υ_+, as always in this thesis. A plane wave travelling in the direction of the positive x-axis is taken to be incident upon the cylinder. The scattered far-field is calculated for a range of

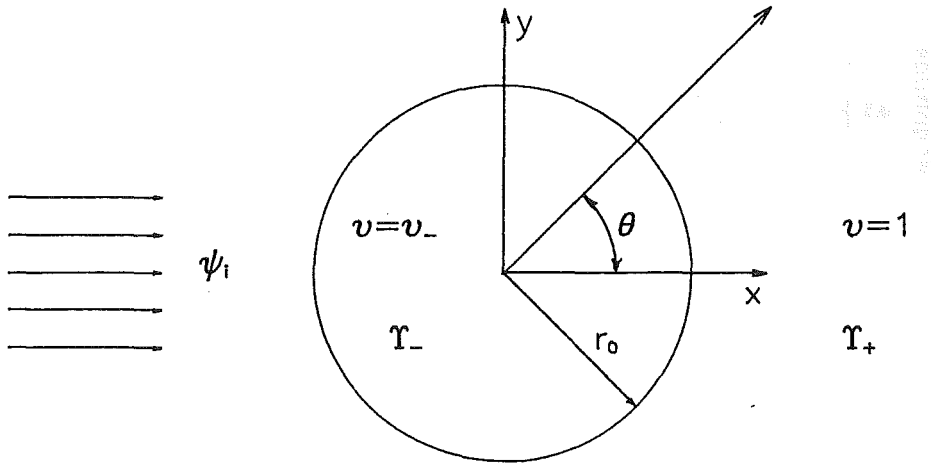


Figure 3.3: The standard scattering object which is a circular cylinder of radius r_0 of constant refractive index. The incident radiation is planar and propagates in the direction of the positive x-axis. The scattered far-field is calculated for a range of scattering angles θ.

scattering angles. Fig. 3.4 illustrates the application of the algorithm introduced above to the calculation of the field scattered from the standard object. The magnitude and phase of the scattered field are plotted verses the scattering angle θ for $\nu_- = 1.01, 1.2$ and 1.5. The results should be compared with those generated by the eigenfunction expansion method (fig. 3.11 and fig. 3.12) which provides exact analytic expressions for the scattering. The results indicate that the accuracy of ray tracing compares well with the analytic solution if the diameter of the scattering object is at least greater than 10λ and the refractive index differs from unity by less than 20 %. These results also show that ray tracing produces a more accurate result than does the Rayleigh-Gans (Born) approximation (fig. 3.9) if $2r_o|\nu_- - 1| > \frac{\lambda}{2}$.

3.1.2 Geometrical Optics

Ray tracing as described in §3.1.1 cannot cope with a medium whose refractive index is discontinuous. This is because the amplitude $A(\mathbf{x})$ is assumed to vary smoothly along the rays whereas the amplitude of the actual field must alter markedly across discontinuities in the refractive index. Consequently, the equations (3.6) and (3.7) upon which ray tracing are based are then inadequate on their own. They need to be supplemented by Geometrical optics (GO) which, although it assumes wave motion can be represented by rays, introduces a refracted ray and a reflected ray whenever a ray is incident upon a point in the medium where there is an abrupt change in the refractive index.

Consider a medium which has a discontinuity in its refractive index across a plane as indicated in fig. 3.5, with the refractive index varying continuously on either side of the plane. Within each continuous part of the medium, ray tracing is performed (invoking the methods described in §3.1.1). Once a ray reaches a discontinuity, refracted

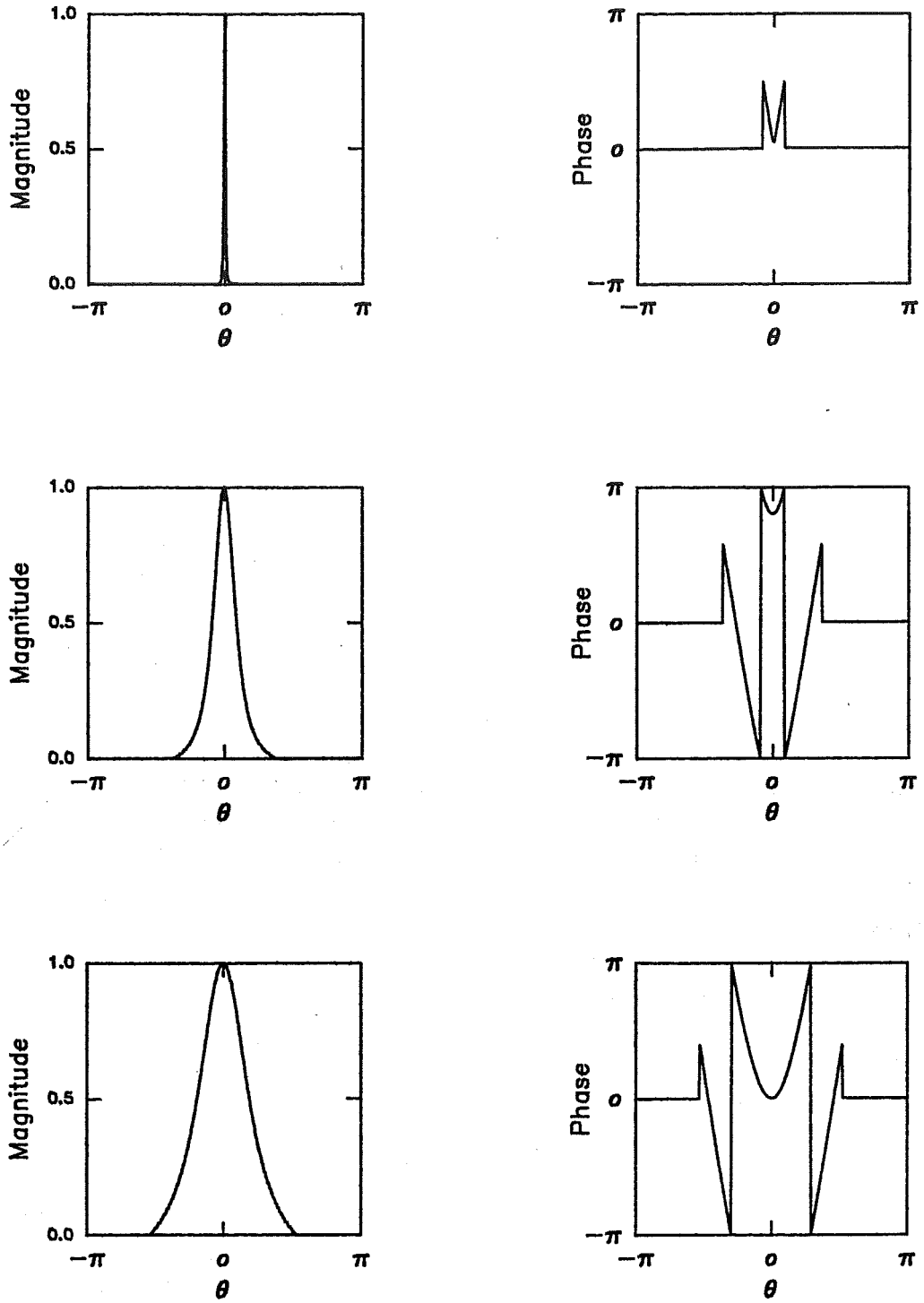


Figure 3.4: The application of the ray tracing algorithm introduced in §3.1.1.1 to the calculation of the magnitude and phase of the far-field scattering from the standard object shown in fig 3.3. The top, middle and bottom row of graphs in the figure correspond, respectively, to the refractive index ν_- of the cylinder being 1.01, 1.2, and 1.5.

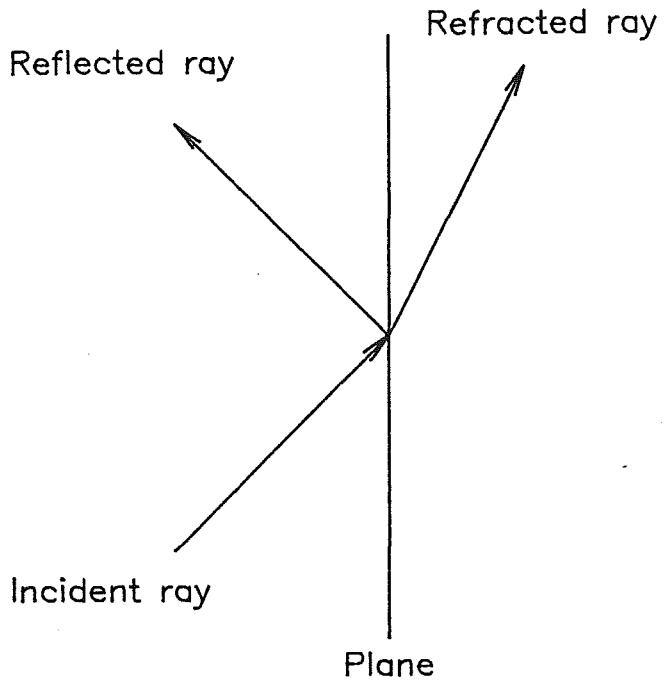


Figure 3.5: Rays in a medium for which the refractive index is discontinuous across a plane. Incident, reflected and refracted rays are shown.

and reflected rays are generated, according to Snell's law of refraction and reflection (James, 1986, Chapter 3). The amplitudes of these rays correspond to the refraction and reflection coefficients for plane waves incident upon planar interfaces, assuming continuity of the total field and its derivative across the boundary (Jones, 1964, Chapter 6; James, 1986, Chapter 2). GO can also cope with a medium whose refractive index is discontinuous along smooth non-planar surfaces. It is assumed that, at each point on such a surface, the field behaves locally as if it is in the vicinity of an infinite planar interface, tangent to the surface at the said point (James, 1986, Chapter 4).

If the medium has more then one surface on which a discontinuity in refractive index occurs, then each discontinuity is treated separately. Each ray is traced individually through the medium until it reaches a discontinuity. At this point the refracted and reflected rays are calculated. These rays are then individually traced until they intersect another discontinuous boundary where the process is repeated. The total field at a point is then simply the superposition of all incident, reflected and refracted rays. The number of multiple interactions to be calculated for each ray depends on the particular problem and the degree of accuracy required.

GO is particularly convenient to apply to a medium whose refractive index is piecewise constant (Keller, 1958; Westcott, 1983). The rays in each piece are then straight. This considerably reduces the computational effort needed for ray tracing. The amplitudes of rays are still calculated using conservation of energy along tubes of rays. However, because the rays are straight, explicit formulas for the amplitudes can be developed, in the following manner. Consider a wave propagating in the direction AA' (see fig. 3.6). Let $d\Sigma_1$ be an element of area on the wave front S_1. Since the wavefront

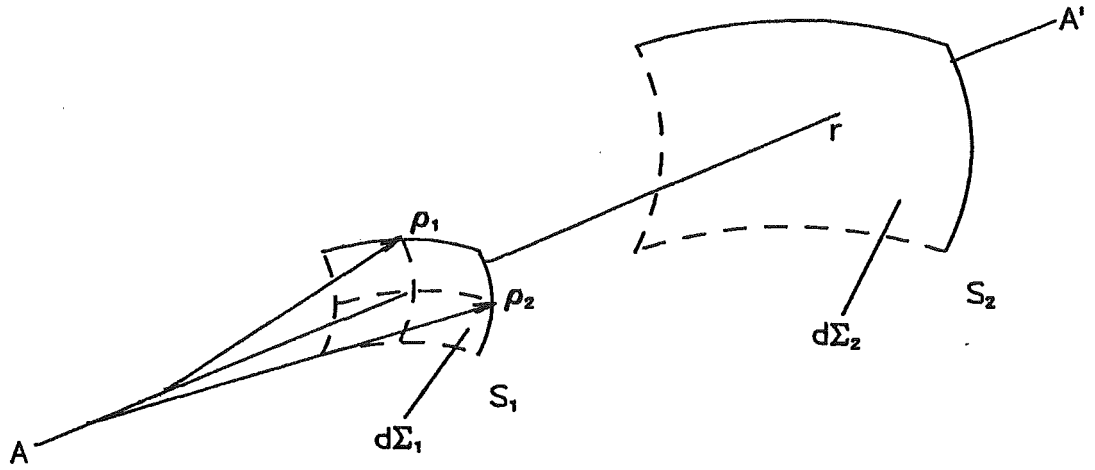


Figure 3.6: Two wavefronts, S_1 and S_2, form a tube of rays in a homogeneous medium. AA' is the axial ray connecting the wavefronts S_1 and S_2.

is in general curved, $d\Sigma_1$ possesses two principle radii of curvature, ρ_1 and ρ_2. The ray, denoted AA' in fig. 3.6, connecting the central points of the elements of area, $d\Sigma_1$ and $d\Sigma_2$ of S_1 and S_2 respectively, is called the axial ray. The rays emanating from each point on $d\Sigma_1$ form a tube of rays about the axial ray which intersect, the element of area $d\Sigma_2$ of the wavefront S_2. Conservation of energy can then be invoked along this tube (as in §3.1.1) to give

$$|A_1|^2 d\Sigma_1 = |A_2|^2 d\Sigma_2 \tag{3.18}$$

where A_1 and A_2 are the complex amplitudes of the wave on $d\Sigma_1$ and $d\Sigma_2$ respectively. The ratio of the two elemental areas is given by

$$\frac{d\Sigma_1}{d\Sigma_2} = \frac{\rho_1 \rho_2}{(\rho_1 + r)(\rho_2 + r)} \tag{3.19}$$

where r is the distance along the axial ray between $d\Sigma_1$ and $d\Sigma_2$. Furthermore, in a homogeneous medium a phase increment of $(-k\nu\, dl)$ results when a ray traverses a distance dl in a medium of constant refractive index ν. Therefore, the complex amplitude of the field A_2, expressed at S_2 in terms of the field A_1 at S_1, is

$$A_2 = A_1 \left[\frac{\rho_1 \rho_2}{(\rho_1 + r)(\rho_2 + r)} \right]^{\frac{1}{2}} e^{-jk\nu r} \tag{3.20}$$

Special forms of the wavefronts can be described by assigning the radii of curvature ρ_1 and ρ_2 particular values. When $\rho_1 = \rho_2$, the field forms a spherical wave. If one of the principle radii is infinite then the field propagates cylindrically. If both principle radii are infinite then the wave is planar.

Provided the fields and the medium are such that high frequency approximations are justified, the only inadequacy of GO is that it cannot handle media possessing boundaries with small radii of curvature. This deficiency is overcome by the Geometrical Theory of Diffraction, which is discussed next.

3.1.3 The Geometrical Theory of Diffraction

The Geometrical Theory of Diffraction (GTD) was developed by Keller (1962) to overcome the inadequacies of GO for media which are discontinuous across boundaries having sharp corners or small radii of curvature. GTD takes account of such discontinuities by supplementing the rays postulated for GO by extra rays, called diffracted rays, which are generated by scattering from sharp edges and sharp corners or whenever incident rays graze smooth boundaries along which the refractive index is discontinuous (Kouyoumjian, 1965). It is convenient to give the name "sharp discontinuities" to the discontinuities with which GTD is specifically concerned, to distinguish them from the "smooth discontinuities" which are the province of GO.

Keller expresses the complex amplitude of each diffracted ray in terms of a diffraction coefficient, denoted by D. The diffraction coefficient relates the amplitude of the ray incident on the sharp discontinuity to the amplitude of the diffracted ray. The diffracted rays propagate away from the discontinuity. In media which are homogeneous, rays propagate according to (3.20). The amplitude A_d of a typical diffracted ray is thus given by

$$A_d = DA_i \left[\frac{\rho_1\rho_2}{(\rho_1 + r)(\rho_2 + r)}\right]^{\frac{1}{2}} e^{-jk\nu r} \tag{3.21}$$

where ν is the constant refractive index of the medium, r is the distance between the sharp discontinuity and the point where the field is observed, A_i is the amplitude of the incident ray, and ρ_1 and ρ_1 are the radii of curvature of the diffracting wavefronts. The curvatures of the latter wavefronts depend on the type of sharp discontinuity and also on the curvatures of the incident wavefronts. For an edge, the wavefronts propagate cylindrically so the term in square brackets reduces to $r^{-\frac{1}{2}}$. For a corner or apex, the wavefronts propagate spherically, so that the term in square brackets reduces to r^{-1}.

The range of directions in which diffracted rays travel is also different for each type of sharp discontinuity. The diffracted rays resulting from a ray incident at a point P on a edge are shown in fig. 3.7. The diffracted rays lie on a cone whose semi-angle corresponds to the angle θ the incident ray makes with the edge. For a corner or apex, the diffracted rays propagate away in all directions. For an incident ray grazing a smooth surface, a single diffracted ray follows the surface, decaying exponentially, with its initial direction equal to that of the incident ray at the grazing point. The rays travel on the surface in such a way that Fermat's principle is satisfied. In other words, the rays trace out geodesics or shortest paths along the surface. All of this is explained in detail by James (1986). The functional forms of the diffraction coefficients are found from asymptotic expansions of exact analytical solutions to certain canonical diffraction problems. These problems are concerned with scattering from objects of a simple shape, such as cylinders, corners and wedges. For each discontinuity the diffraction coefficient is found by subtracting from the exact solution to the canonical problem the GO approximation to the same problem (Kouyoumjian, 1965; Bowman *et al.*, 1969; James, 1986; Herrmann, 1987).

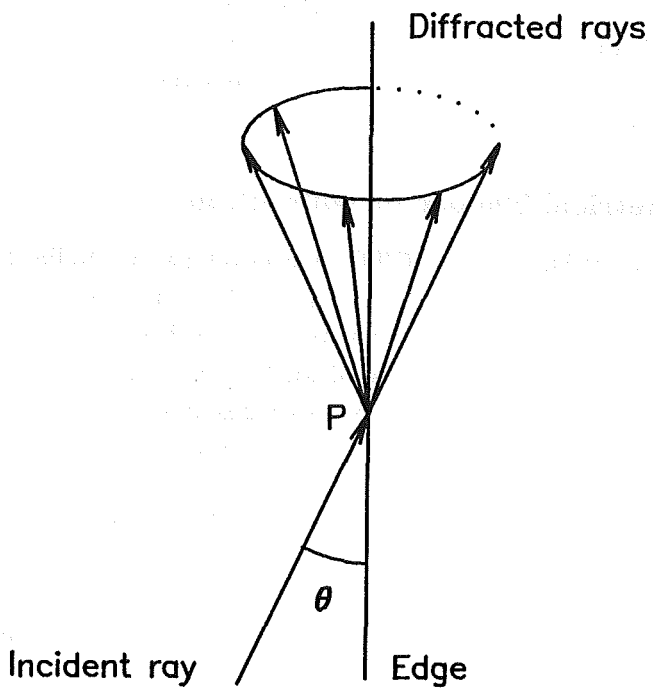


Figure 3.7: Diffraction of a ray incident at a point P on a edge. The diffracted rays make a cone whose semi-angle corresponds to the angle θ the incident ray makes with the edge.

In the real-world the sharp discontinuities are generally of a finite size. GTD can still be invoked however, by assuming that the field in the neighbourhood of each part of a discontinuity behaves locally as if it relates to one of the canonical problems. The total diffracted field is then obtained by vectorially summing the rays from each part of each individual discontinuity on the scatterer. Improvements to the approximation can be obtained by letting the diffracted rays from one discontinuity be incident on another. The total solution then represents the incorporation of higher order diffraction components.

Throughout the last quarter of a century GTD has proved valuable for solving a wide variety of electromagnetic problems to an impressive level of accuracy. One of the earliest examples is Kinber (1962) calculation of the radiation from a horn antenna. GTD is now "standard practice", as emphasised by many authors (cf Cheriot, 1971; Ufimtsev, 1971; Kouyoumjian and Pathak, 1974; Lee and Deschamps, 1976; Mittra *et al.*, 1979; James, 1986)

It is worth noting that closed-form expressions for diffraction coefficients for penetrable sharp discontinuities have not yet been developed, because of the apparently intractable nature of the corresponding canonical problems, such as diffraction by a dielectric prism or penetrable acoustic wedge (Berntsen, 1983; Yeo, 1985).

3.1.4 The WKB Method

The WKB method generates asymptotic solutions to differential equations of the general form

$$\left(p_1(\mathbf{x})\nabla^2 + kp_2(\mathbf{x})\nabla + k^2 q_1(\mathbf{x}) + kq_2(\mathbf{x}) + q_3(\mathbf{x})\right)\psi(\mathbf{x},k) = 0 \qquad (3.22)$$

when k tends to infinity and the $p_1(\mathbf{x})$, $p_2(\mathbf{x})$, $q_1(\mathbf{x})$, $q_2(\mathbf{x})$, $q_3(\mathbf{x})$ are real functions of position $\mathbf{x}$ (Jones, 1964, Chapter 6). The WKB method is named after Wentzel, Kramers and Brillouin. However Liouville, Gans, Jefferys and Langer are also associated with the development of the method (Wait, 1962, Chapter 4). The WKB method is useful for direct problems at high frequencies because the Helmholtz equation (2.3) is a specialised form of (3.22). It can also be useful for obtaining asymptotic expansions of higher functions of analysis (such as Bessel functions), satisfying differential equations which are other specialisations of (3.22) (Olver, 1974).

The WKB method can be regarded as a refined form of ray tracing. An expansion of the form

$$\psi(\mathbf{x},k) = e^{-jkS(\mathbf{x})}\sum_{n=0}^{\infty} A_n(\mathbf{x})(jk)^{-n} \qquad (3.23)$$

is invoked to represent the field at high frequencies, with the $A_n(\mathbf{x})$ being independent of k. It transpires that $A_0(\mathbf{x})$ characterises ray tracing as introduced in §3.1.

By substituting (3.23) into the Helmholtz equation (2.3), and separately equating to zero the coefficient of each power of k, the following series of equations is obtained,

$$\nabla S \cdot \nabla S = \nu^2 \qquad (3.24)$$

$$\nabla^2 S A_0 + 2\nabla S \cdot \nabla A_0 = 0 \qquad (3.25)$$

and, for $n > 0$,

$$A_{n+1}\nabla^2 S + 2\nabla S \cdot \nabla A_{n+1} + j\nabla^2 A_n = 0 \qquad (3.26)$$

where (3.24) is recognised as the eikonal equation, previously stated in (3.6). The relations (3.25) and (3.26) are called transport equations. The most significant term of (3.23) is usually A_0, but the higher order terms can become important wherever the scale of spatial variations of the refractive index are of the order of a wavelength. The equations (3.25) and (3.26) allow the unknowns in (3.23) to be determined giving an asymptotic solution to (3.22).

The WKB method is employed in numerous sciences, well known examples of which are atmospheric propagation, quantum mechanics and wave scattering in layered media (Heading, 1962; Wait, 1962; Jones, 1964; Born and Wolf, 1970).

3.2 Volume Source Formulation

The volume source formulation allows a general implicit solution to (2.3) to be written down immediately. This formulation can form the basis of several exact and approximate solutions to the direct problem. It can also give conceptual insight into scattering phenomena.

The volume source formulation is based on (2.7) which is rewritten here as

$$\left(\nabla^2 + k^2\right)\psi_s(\mathbf{x},k) = -\tilde{\sigma}(\mathbf{x},k) = -k^2\left[\chi(\mathbf{x},k)^2 - 1\right]\left[\psi_i(\mathbf{x},k) + \psi_s(\mathbf{x},k)\right] \qquad (3.27)$$

where $\psi_i(\mathbf{x}, k)$ is understood to satisfy (2.6) and where the significance of the symbol $\tilde{\sigma}(\mathbf{x}, k)$ is explained below. Both LHS (3.27) and LHS (2.6) imply that wave motion travels everywhere at the free space wave speed. Although this contradicts the known behaviour of wave motion in refracting media (Jones, 1964, Chapter 6; Tan, 1985) it involves no actual approximation because equivalent reradiating sources, of volume density $\tilde{\sigma}(\mathbf{x}, k)$, are introduced wherever the propagation medium differs from free space. These sources are also known as polarisation sources since they represent the polarisation of the medium by the field incident upon it (Bates and Ng, 1972).

By superposition, the Green's function (see §2.2.4) can be invoked to develop an integral representation for the field from the source density $\tilde{\sigma}(\mathbf{x}, k)$. At an arbitrary point $\mathbf{x}$, the contribution from each source element is summed. The field at $\mathbf{x}$ due to a source density $\tilde{\sigma}(\mathbf{x}, k)$ is then the integral of all contributions over the volume Υ_- occupied by the sources.

Consequently, the formal solution to the direct problem in the volume source formulation is

$$\begin{aligned} \psi_s(\mathbf{x}, k) &= \int_{\Upsilon_-} \tilde{\sigma}(\mathbf{x}_1, k) g(\mathbf{x}, \mathbf{x}_1, k)\, d\Upsilon(\mathbf{x}_1) \\ &= \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x}_1, k) - 1)\psi(\mathbf{x}_1, k) g(\mathbf{x}, \mathbf{x}_1, k)\, d\Upsilon(\mathbf{x}_1) \end{aligned} \tag{3.28}$$

Because it is only the scattered field which is reradiated by the volume source density $\tilde{\sigma}(\mathbf{x}, k)$, the total field is obtained by adding the incident field to (3.28), i.e.

$$\psi(\mathbf{x}, k) = \psi_i(\mathbf{x}, k) + \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x}_1, k) - 1)\psi(\mathbf{x}_1, k) g(\mathbf{x}, \mathbf{x}_1, k)\, d\Upsilon(\mathbf{x}_1) \tag{3.29}$$

The version (3.29) of the volume source formulation is a Fredholm integral equation of the second kind (see §2.2.8), which can be solved numerically for $\psi(\mathbf{x}, k)$ given $\psi_i(\mathbf{x}, k)$ and $\chi(\mathbf{x}, k)$. When the volume of Υ_-, in cubic wavelengths, exceeds what can be handled by the available computational facilities, it is necessary to resort to approximate methods, such as those due to Rayleigh-Gans (Born) and Rytov, in order to generate numerical solutions to either (3.28) or (3.29).

3.2.1 Rayleigh-Gans (Born) Approximation

This approximation, which is due to Lord Rayleigh, was further developed by Gans. It was adapted by Born to non-relativistic quantum mechanical contexts to which it is particularly suited because the refractive index is effectively unity everywhere for the Schrödinger equation.

The Rayleigh-Gans (Born) approximation is predicated on approximating the total field by the incident field inside the region Υ_-. The volume source formulation (3.28) then becomes

$$\psi_b(\mathbf{x}, k) = \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x}_1, k) - 1)\psi_i(\mathbf{x}_1, k) g(\mathbf{x}, \mathbf{x}_1, k) d\Upsilon(\mathbf{x}_1) \tag{3.30}$$

where $\psi_b(\mathbf{x}, k)$ is the Rayleigh-Gans (Born) approximation to the scattered field. Since all the quantities in the integrand are known, (3.30) is an explicit integral expression for the scattered field. This allows computationally efficient algorithms to be developed for solving the direct problem.

It is however important to understand for which types of media the Rayleigh-Gans (Born) approximation gives useful results. The concept of wavefronts introduced in §3.1.1 is central to this. The incident field can be expected to be a useful approximation to the total field if the latter's wavefronts do not deviate appreciably from the incident wavefronts. An appreciable deviation can be said to occur when the total path length of a ray passing through Υ_- deviates from that of the incident ray by more than $\lambda/2$ (Goodman, 1968). Consider any incident ray, of length L say, transversing Υ_-. Inspection of (3.10) shows that, while the path length of this incident ray is L, the path length of the actual ray corresponding most closely to the incident ray is approximately $\int_{l_1}^{l_2} \nu(\mathbf{x})\, dl$, where l_1 and l_2 identify the points where the incident ray enters and leaves Υ_- and dl is the element of length along the ray. Consequently, the Rayleigh-Gans (Born) approximation can be expected to be valid if the refractive index satisfies $|\int_{l_1}^{l_2} \nu(\mathbf{x})\, dl - L| < \lambda/2$ for all incident rays transversing Υ_-.

When $\psi_i(\mathbf{x}, k)$ is a plane wave, and $\psi_s(\mathbf{x}, k)$ is evaluated in the far field, the Rayleigh-Gans (Born) approximation reduces to the Fourier transform of $k^2[\chi^2(\mathbf{x}, k)-1]$ (Jones, 1964, Chapter 6). To show this, one substitutes the expression (3.2) for $\psi_i(\mathbf{x}, k)$ and the far-field form (2.19) for the Green's function into (3.30). The Rayleigh-Gans (Born) approximation then becomes

$$\psi_{ff_b}(\hat{\mathbf{x}}, k) = \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x}_1, k) - 1) e^{-jk(\hat{\mathbf{k}}-\hat{\mathbf{x}})\cdot\mathbf{x}_1} d\Upsilon(\mathbf{x}_1) \tag{3.31}$$

where $\psi_{ff_b}(\hat{\mathbf{x}}, k)$ is the Rayleigh-Gans (Born) approximation to $\psi_s(\mathbf{x}, k)$ evaluated in the far-field (see §2.2.4). On defining $\mathbf{k}' = k(\hat{\mathbf{k}} - \hat{\mathbf{x}})$ and invoking the definition of the Fourier transform given in §2.2.6, the Rayleigh-Gans (Born) approximation can be written as

$$\psi_{ff_b}(\hat{\mathbf{x}}, \mathbf{k}) = k^2 \mathcal{F}_2 \left[\chi^2(\mathbf{x}) - 1\right](\mathbf{k}') \tag{3.32}$$

If the scattered field is observed for all $\hat{\mathbf{x}}$ and the incident field is a plane wave whose direction of incidence is $\hat{\mathbf{k}}$, then the extremity of the vector $\mathbf{k}' = k(\hat{\mathbf{k}} - \hat{\mathbf{x}})$ traces out a circle, as illustrated in fig. 3.8. The radius of this circle is k and its origin is located at the extremity of $\mathbf{k}$. Consequently, the Fourier transform of $k^2(\chi^2(\mathbf{x}) - 1)$ on this circle, represents the observable scattered field due to a incident plane wave $e^{-j\mathbf{k}\cdot\mathbf{x}}$. The Rayleigh-Gans (Born) approximation can also be thought of as the first term of a series converging to the exact solution to (3.28) under appropriate conditions (Arfken, 1985, Chapter 16). This series can be formally constructed in terms of the integral operator

$$[K\psi](\mathbf{x}, k) = \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x}_1, k) - 1)\psi(\mathbf{x}_1, k) g(\mathbf{x}, \mathbf{x}_1, k) d\Upsilon(\mathbf{x}_1) \tag{3.33}$$

The series can be written as

$$\psi_s(\mathbf{x}, k) = \lim_{n\to\infty} \sum_{i=1}^{n} \psi^{(i)}(\mathbf{x}, k) = \lim_{n\to\infty} \sum_{i=0}^{n-1} K\psi^{(i)}(\mathbf{x}, k) \tag{3.34}$$

where the terms are defined by

$$\begin{aligned} \psi^{(0)}(\mathbf{x}, k) &= \psi_i(\mathbf{x}, k) \\ \psi^{(1)}(\mathbf{x}, k) &= K\left[\psi^{(0)}\right](\mathbf{x}, k) \\ \psi^{(i)}(\mathbf{x}, k) &= K\left[\psi^{(i-1)}\right](\mathbf{x}, k) \\ &= K^i \psi^{(0)}(\mathbf{x}, k) \end{aligned} \tag{3.35}$$

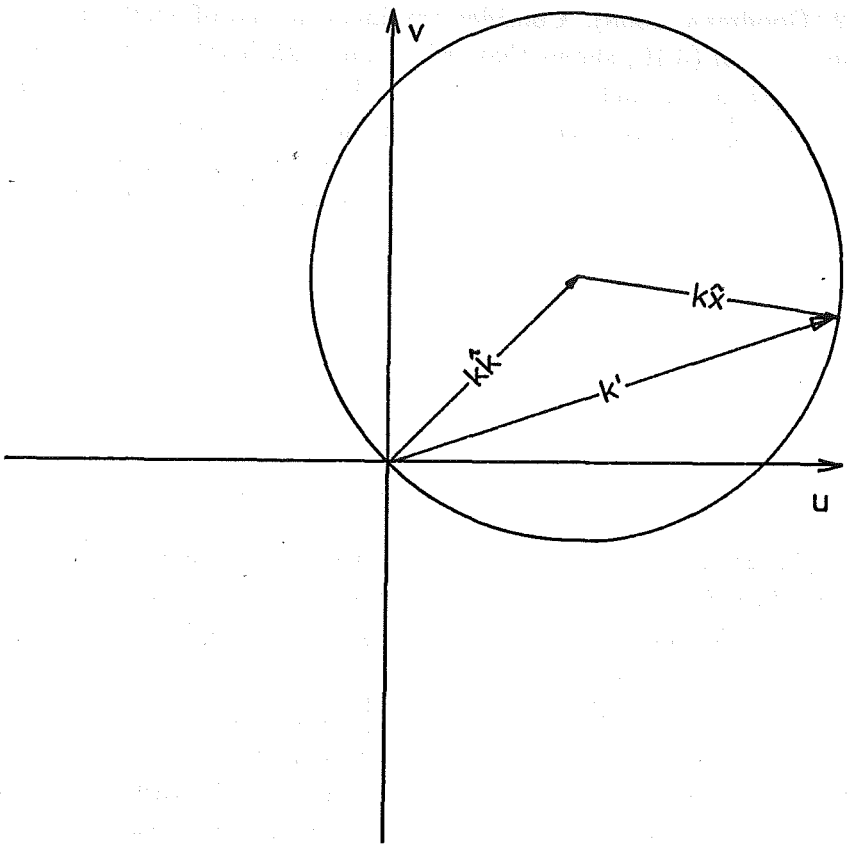


Figure 3.8: Locus in Fourier space corresponding to observable scattered field. The scattered field is related to the parts of the Fourier transform of $k^2(\chi^2(\mathbf{x}) - 1)$ on a circle in Fourier space. The centre of the circle lies at the extremity of the vector $\mathbf{k}$, where $\mathbf{k}$ defines the direction in which the incident field travels.

The series (3.34) is the unique solution to the direct problem, provided the limit on the RHS (3.34) exists. The sequence of terms in (3.34) constitutes what has become known as the Rayleigh-Gans (Born) series. Convergence of this series is governed by the sufficient but not necessary condition $\| K \| < 1$, where $\| \cdot \|$ is some appropriate norm. This condition on the norm of the operator K is equivalent to (Leeman *et al.*, 1985)

$$k^2 \max_{\mathbf{x} \in \Upsilon_-} |\chi^2(\mathbf{x}, k) - 1| L^3 < 1 \tag{3.36}$$

where L is the length of the longest incident ray transversing Υ_-.

Because the Rayleigh-Gans (Born) approximation is invoked repeatedly in this thesis, both directly and by comparison with other techniques, an example of its use is now presented. Fig. 3.9 illustrates the application of (3.32) to the calculation of the field scattered from the standard object (see fig. 3.3). The magnitude and phase of the scattered field is shown in fig. 3.9 for $2r_o = 1.25\lambda$, $2r_o = 5\lambda$ and $2r_o = 10\lambda$. The results should be compared with those generated by the eigenfunction expansion method (fig. 3.11 and fig. 3.12) which provides exact analytic expressions for the scattering. The results indicate that the accuracy of the Rayleigh Gans (Born) approximation compares well with the analytic solution if $2r_o|\nu_- - 1| < \frac{\lambda}{2}$. If the product $2r_o|\nu_- - 1|$ becomes greater than $\frac{\lambda}{2}$, proportionally larger percentage errors between the actual and approximate scattered field must be expected.

3.2.2 The Rytov Approximation

Although the Rayleigh-Gans (Born) approximation is very convenient, it is only valid to approximate $\psi(\mathbf{x}, k)$ within Υ_- by $\psi_i(\mathbf{x}, k)$ under highly restricted sets of circumstances, as intimated in §3.2.1. An approximation which could incorporate wavefront curvature as the field propagates through a medium would yield a more realistic expression for the total field. To this end, therefore, let the total field be described by

$$\psi(\mathbf{x}, k) = e^{-jk(S_i(\mathbf{x}) + S_s(\mathbf{x}))} \tag{3.37}$$

where $S_s(\mathbf{x})$ is a complex function of position and $S_i(\mathbf{x})$ is defined by

$$\psi_i(\mathbf{x}, k) = e^{-jkS_i(\mathbf{x})} \tag{3.38}$$

The real and imaginary parts of $S_i(\mathbf{x}) + S_s(\mathbf{x})$ can be considered to incorporate wavefront curvature and amplitude variations, respectively, of the field as it propagates through the medium. The form of (3.37) can be compared to the approximate expression (3.4) for the field at high frequencies.

Taking $\psi_i(\mathbf{x}, k)$ to satisfy the free-space wave equation shows that $S_i(\mathbf{x})$ obeys

$$\nabla^2 S_i - jk(\nabla S_i \cdot \nabla S_i - 1) = 0 \tag{3.39}$$

Using (2.3) shows that $S_s(\mathbf{x})$ satisfies

$$\nabla^2 S_s - jk\left[\nabla S_s \cdot \nabla S_s + 2\nabla S_i \cdot \nabla S_s - (\chi 2(\mathbf{x}) - 1)\right] = 0 \tag{3.40}$$

The differential equations (3.39) and (3.40) are intractable as they stand. It is necessary to introduce some approximation in order to make any useful headway.

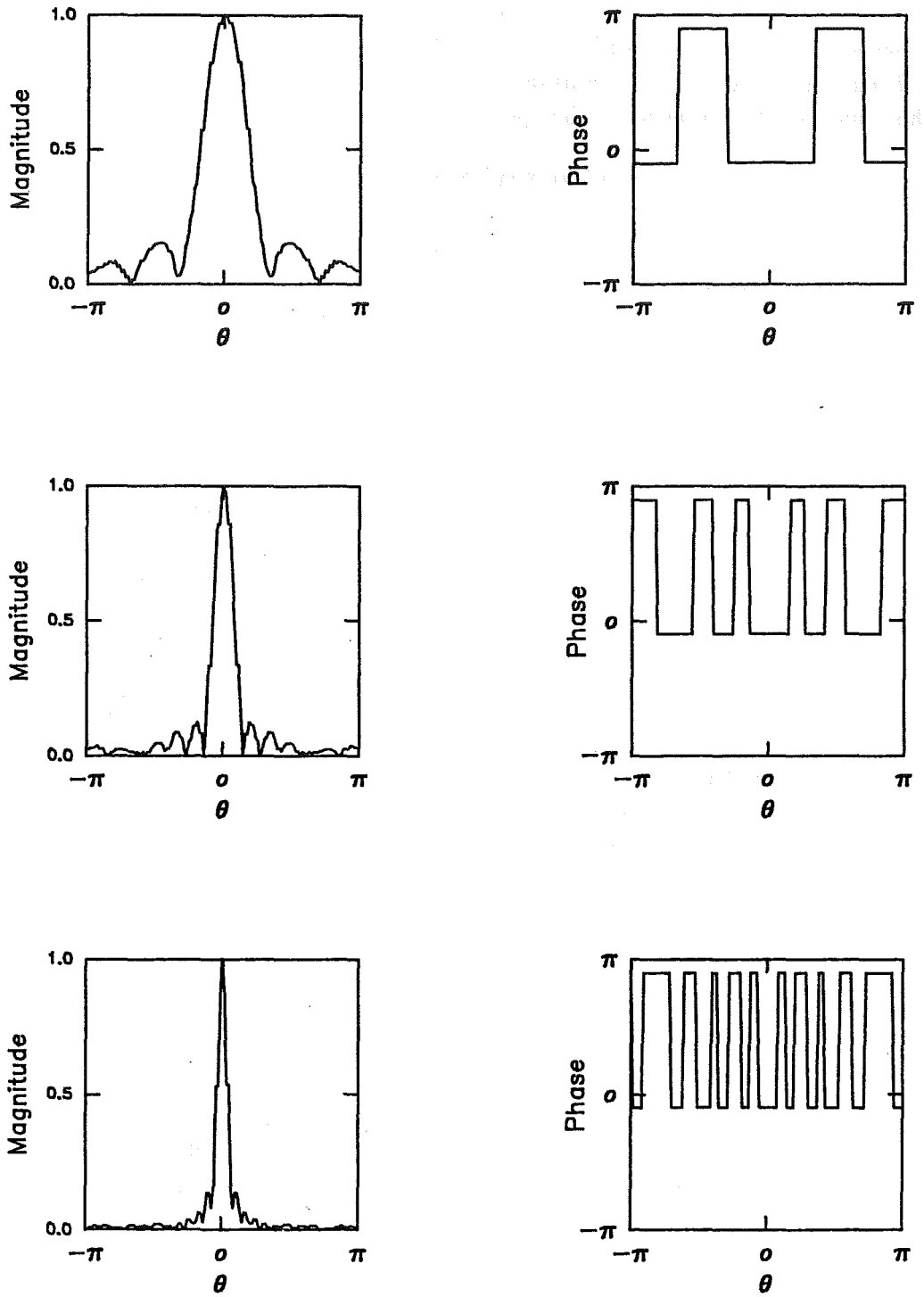


Figure 3.9: Illustration of application of the Rayleigh-Gans (Born) approximation. The magnitude and phase of the field scattered from the standard object (shown in fig 3.3) are calculated using (3.32). The top, middle and bottom row of graphs in the figure correspond, respectively, to the scattering object for $2r_o = 1.25\lambda$, $2r_o = 5\lambda$ and $2r_o = 10\lambda$.

On introducing Rytov's transformation (Chernov, 1967)

$$\tilde{\psi}(\mathbf{x},k) = -jk\, S_s(\mathbf{x}) e^{-jkS_i(\mathbf{x})} \tag{3.41}$$

it can be seen that

$$\nabla^2\tilde{\psi}(\mathbf{x},k) + k^2\tilde{\psi}(\mathbf{x},k) = -k^2\left[(\chi^2(\mathbf{x})-1) - \nabla S_s \cdot \nabla S_s\right]\psi_i(\mathbf{x},k) \tag{3.42}$$

which is the familiar form of the Helmholtz equation with sources, i.e. it is equivalent to (2.7). The only term on RHS (3.42) which is not given for the direct problem is $\nabla S_s \cdot \nabla S_s$. The Rytov approximation is obtained by neglecting $\nabla S_S \cdot \nabla S_s$ in (3.42), i.e. (Chernov, 1967)

$$\nabla^2\tilde{\psi}(\mathbf{x},k) + k^2\tilde{\psi}(\mathbf{x},k) = -k^2(\chi^2(\mathbf{x})-1)\psi_i(\mathbf{x},k) \tag{3.43}$$

Since RHS (3.43) depends only on the incident field, it may be transformed to an integral expression similar to (3.30). This integral expression may then be converted to a Fourier integral under the conditions described in §3.2.1. However, it is necessary to discard $\psi_s(\mathbf{x},k)$ and replace it with $\tilde{\psi}(\mathbf{x},k)$, which can be done by substituting (3.37) and (3.38) into (3.41). Thus

$$\tilde{\psi}(\mathbf{x},k) = \psi_i(\mathbf{x},k)\left[\ln\psi(\mathbf{x},k) - \ln\psi_i(\mathbf{x},k)\right] \tag{3.44}$$

The advantage of the Rytov approximation over the Rayleigh-Gans (Born) approximation is that it can be valid for a wider range of media. It may seem true in general that relative amplitude and incremental phase changes (as defined in §2.2.3) of $\psi(\mathbf{x},k)$ only need to be small per wavelength, in contrast to the Rayleigh-Gans (Born) approximation for which the incremental phase changes of $\psi(\mathbf{x},k)$ must be small across the entire medium. For this reason it is claimed by Chernov (1967) that the Rytov approximation is more accurate for extended scatterers than the Rayleigh-Gans (Born) approximation. This is only partly true, however, because Keller (1969) has shown that the Rytov approximation is only superior to the Rayleigh-Gans (Born) approximation for calculating scattering which is predominantly in the forward direction (i.e. in the direction of the incident field, which is understood to be a plane wave, or at least a close approximation to one).

Bates *et al.* (1976) introduce an extension to the Rytov approximation which further increases its range of validity. The term $\nabla S_s \cdot \nabla S_s$ in (3.42) is partially compensated so that its distorting effects are somewhat reduced. Furthermore, this extended Rytov approximation is as convenient for inverse scattering as the Born approximation.

3.3 The Null-Field Method

The null-field method provides an exact integral representation of the field scattered by homogeneous and impenetrable objects (Bates and Wall, 1977). The method is based upon an extension of the idea that fields cannot penetrate a totally reflecting object. This idea, due originally to Ewald and Oseen (1915), is now called the Ewald-Oseen extinction theorem (Agarwal, 1976). It has been formalised by Waterman (1965) as the extended boundary condition. Further development of these ideas has allowed similar treatments of penetrable homogeneous objects to be devised (Waterman, 1969; Peterson and Ström, 1975; Varatharajulu and Pao, 1976; Varadan *et al.*, 1978).

The Ewald-Oseen extinction theorem states that the scattered field extinguishes (i.e. exactly cancels) the incident field throughout Υ_-. For a totally reflecting body it is physically obvious that a field cannot penetrate inside Υ_-. However, even if a body is penetrable, it is possible to reformulate the problem in such a way that the scattered field remains extinguished within Υ_- (Bates, 1980).

The null-field method satisfies the extinction theorem by employing equivalent sources on the surface of the scattering object. These sources radiate such that the field is extinguished in Υ_-. To calculate the field radiating from sources on a surface it is convenient to invoke the Huygens-Kirchhoff theorem (Goodman, 1975), which states that the field $\psi(\mathbf{x}, k)$, at an arbitrary point in a homogeneous medium, can be expressed in terms of the field and its normal derivative everywhere on an arbitrary closed surface Γ surrounding the point. So, (Morse and Feshbach, 1953, Chapter 7)

$$h(\mathbf{x}, \nu) = \int_\Gamma \left[\psi(\mathbf{x}_1, k) \frac{\partial g(\mathbf{x}, \mathbf{x}_1, k\nu)}{\partial \mathbf{n}} - g(\mathbf{x}, \mathbf{x}_1, k\nu) \frac{\partial \psi(\mathbf{x}_1, k)}{\partial \mathbf{n}} \right] d\Gamma(\mathbf{x}_1) \tag{3.45}$$

where $h(\mathbf{x}, \nu) = \psi(\mathbf{x}, k)$, $\mathbf{n}$ is the outward normal to $\Gamma(\mathbf{x})$ and $\psi(\mathbf{x}, k)$ and $\frac{\partial \psi(\mathbf{x},k)}{\partial \mathbf{n}}$ are known as the equivalent surface sources.

If a homogeneous scattering region is characterised by

$$\nu(\mathbf{x} \in \Upsilon_-) = \nu_- \tag{3.46}$$

where ν_- is a constant then the Huygens-Kirchhoff theorem can be invoked for the surface $\Gamma(\mathbf{x}) = \sigma(\mathbf{x})$ to give (Baker and Copson, 1950, Chapter 1; Bates, 1980)

$$\nu = 1 \quad x \in \Upsilon_+ \quad h(\mathbf{x}, \nu) = \psi_s(\mathbf{x}, k) \tag{3.47}$$

$$\nu = \nu_- \quad x \in \Upsilon_+ \quad h(\mathbf{x}, \nu) = \psi(\mathbf{x}, k) \tag{3.48}$$

$$\nu = 1 \quad \mathbf{x} \in \Upsilon_- \quad h(\mathbf{x}, \nu) = -\psi_i(\mathbf{x}, k) \tag{3.49}$$

$$\nu = \nu_- \quad \mathbf{x} \in \Upsilon_+ \quad h(\mathbf{x}, \nu) = 0 \tag{3.50}$$

Integral equations for the quantities $\frac{\partial \psi}{\partial \mathbf{n}}$ and ψ can be formulated by invoking (3.49) and (3.50). Integral representations for the scattered field in Υ_- and Υ_+ can then be written down with the aid of (3.47) and (3.48).

Special cases of the equations (3.47) through (3.50) exist for objects which have Dirichlet or Neumann boundary conditions, which are respectively when $\psi(\mathbf{x}, k) = 0$ and $\frac{\partial \psi(\mathbf{x},k)}{\partial \mathbf{n}} = 0$ on $\sigma(\mathbf{x})$. These conditions are also known as sound soft and sound hard respectively. With either of these boundary conditions the null-field method is simplified since (3.48) and (3.50) are not needed. Either Dirichlet or Neumann conditions are assumed throughout the treatment of one of the implicit exact inverse methods presented in Chapter 5, in which (for the reasons given in §1.5) all the illustrative numerical examples are two-dimensional. The rest of this section is devoted to developing the two-dimensional null-field formulas.

Refer to §1.3.1 for the adaptation to two dimensions of the three-dimensional notation employed above. The points $\mathbf{x}$ and $\mathbf{x}_1$ are accorded the cylindrical polar coordinates (ρ, ϕ) and (r, θ) respectively. It is adequate, for the purposes of the illustrative examples invoked in this thesis, to take the closed curve $\sigma(\mathbf{x}_1)$ to be star-like (i.e. single valued with respect to θ) so that it can be parameterised as

$$\sigma(\mathbf{x}_1) = \sigma(\theta) \tag{3.51}$$

The element length $d\sigma(\theta)$ can then be expressed in terms of $d\theta$ as

$$d\sigma(\theta) = \sqrt{\sigma^2(\theta) + (d\sigma/d\theta)^2}\, d\theta \tag{3.52}$$

Now consider the Dirichlet boundary condition, for which (3.45) reduces to

$$h(\mathbf{x}, \nu) = \int_\sigma \frac{\partial \psi(r; \theta, k)}{\partial \mathbf{n}} g(\rho; \phi, r; \theta, k\nu) d\sigma(\theta) \tag{3.53}$$

Invoking (3.49), and substituting (2.17) and (2.11) into (3.53), gives

$$\sum_{m=-\infty}^{\infty} a_m J_m(k\rho) e^{jm\phi} = \frac{-j}{4} \int_\sigma \frac{\partial \psi}{\partial \mathbf{n}} \sum_{m=-\infty}^{\infty} H_m^{(2)}(kr) J_m(k\rho) e^{jm(\phi-\theta)} d\sigma(\theta) \tag{3.54}$$

which must be satisfied throughout Υ_-. The orthogonality of the $e^{jm\phi}$ functions is now invoked within Υ_-, wherein ϕ spans 0 to 2π. Multiplying (3.54) through by $e^{-jm\phi}$ and integrating with respect to ϕ gives, for all integers m,

$$C_m \int_0^{2\pi} f(\theta) H_m^{(2)}(k\sigma(\theta))\, e^{-jm\theta} d\theta = U_m, \quad m = -\infty, \ldots, \infty, \tag{3.55}$$

where $U_m = 4jC_m a_m$, with C_m being weight functions which are multiplied to both sides in order to achieve numerical stability (as explained later). The surface source density $f(\theta)$ is defined to be

$$f(\theta) = \sqrt{\sigma^2(\theta) + (d\sigma/d\theta)^2}\, \frac{\partial \psi(\sigma(\theta), \theta, k)}{\partial \mathbf{n}} \tag{3.56}$$

since the element of length $d\sigma(\theta)$ is related to $d\theta$ through (3.52). Analytic continuation automatically ensures that the incident field is extinguished throughout Υ_- (Bates, 1969). The equations belonging to the set (3.55) are called the null-field equations for the Dirichlet boundary conditions (Bates, 1969; Bates and Wall, 1977).

Equation (3.55) specifies an infinite set of non-singular integral equations for the unknown source density $f(\theta)$. These equations can be solved by invoking the method of moments (Harrington, 1968), which transforms (3.55) into a finite set of linear, algebraic equations. This is effected by expanding $f(\theta)$ as

$$f(\theta) = \sum_{m=-\infty}^{\infty} f_m e^{jm\theta}, \tag{3.57}$$

The null-field equations (3.55) then reduce to the infinite system of linear equations

$$\sum_{n=-\infty}^{\infty} H_{mn} f_n = U_m, \quad m = -\infty, \ldots, \infty, \tag{3.58}$$

where, for $m, n = -\infty, \ldots, \infty$,

$$H_{mn} = C_m \int_0^{2\pi} H_m^{(2)}(k\sigma(\theta)) e^{j(n-m)\theta}\, d\theta \tag{3.59}$$

Once the surface source density has been calculated, the scattered field in Υ_+ can be determined by invoking (3.47) and (3.53). The scattered field can consequently be written as

$$\begin{aligned}\psi_s(\rho;\phi,k) &= \frac{-j}{4}\int_0^{2\pi} f(\theta)\sum_{m=-\infty}^{\infty} H_m^{(2)}(k\rho) \\ &\quad J_m(k\sigma(\theta))e^{jm(\theta-\phi)}d\theta \end{aligned} \tag{3.60}$$

Substituting the series expression for the source density (3.57) into (3.60) results in

$$\psi_s(\rho;\phi,k) = \sum_{m=-\infty}^{\infty} b_m H_m^{(2)}(k\rho)e^{jm\phi} \tag{3.61}$$

where the b_m are given by

$$b_m = \sum_{n=-\infty}^{\infty} J_{mn} f_n, \tag{3.62}$$

with

$$J_{mn} = \int_0^{2\pi} J_m(k\sigma(\phi))e^{j(n-m)\phi}\,d\phi. \tag{3.63}$$

In the far-field (refer to §2.2.2) the Hankel function in (3.61) can, by definition, be approximated by its asymptotic form (2.18) so that the scattered field is given by

$$\psi_s \sim (1/8\pi kr)^{-1/2}e^{-j(kr-3\pi/4)}\sum_{m=-\infty}^{\infty} b_m j^m e^{jm\phi}, \tag{3.64}$$

Note that, since $J_m(\cdot) = \mathrm{Re}(H_m^{(2)}(\cdot))$, the integral in (3.63) has already been explicitly evaluated, for each m, once the H_{mn}, defined by (3.59), have been computed.

When solving the direct problem for a particular $\sigma(\mathbf{x})$, the conditioning of the linear system (3.58) is improved by suitable normalisation of the Hankel functions appearing in (3.59) (Wall, 1980). Equivalently, this may be accomplished by choosing the constants C_m to be $\frac{1}{H_m^{(2)}}(k\sigma(0))$. Thus, in the event that $\sigma(\theta)$ is a circle of radius r_0, H_{mn} reduces to the Kronecker delta, and (Wall *et al.*, 1985)

$$\begin{aligned} f_m &= (-j)^m/H_m^{(2)}(kr_0), \\ b_m &= (-j)^m J_m(kr_0)/H_m^{(2)}(kr_0) \end{aligned} \tag{3.65}$$

as can be deduced from separation of variables (Bowman *et al.*, 1969). It is worth remarking that the numerical stability of the null-field method improves as the ratio of the area in Υ_{--} to that in Υ_- approaches unity (Lewin, 1970).

The scattered fields from objects which satisfy the Neumann boundary condition can be determined similarly. The main difference between the Dirichlet and Neumann cases being that $\frac{\partial g}{\partial \mathbf{n}}$, rather than g, has to be expanded in a multipole series. Note that $\frac{\partial g}{\partial \mathbf{n}}$ can be expressed as $\nabla g \cdot \hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit normal to $\sigma(\mathbf{x})$. The unit normal $\hat{\mathbf{n}}$ on $\sigma(\mathbf{x})$ at the point $\mathbf{x}_1$ is (Murch *et al.*, 1988a),

$$\hat{\mathbf{n}} = \left(\sigma(\theta)\hat{\mathbf{r}} - \frac{d\sigma(\theta)}{d\theta}\hat{\theta}\right)/\sqrt{\sigma^2(\theta) + (d\sigma(\theta)/d\theta)^2} \tag{3.66}$$

where $\hat{\mathbf{r}}$ and $\hat{\theta}$ are the unit cylindrical polar coordinate vectors (Arfken, 1985, Chapter 2). The gradient of the Green's function at the point $\mathbf{x}$ is obtained by taking the gradient of (2.17) to give

$$\nabla g = \sum_{-\infty}^{\infty} \left[k H_m^{(2)\prime}(kr)\hat{\mathbf{r}} - \frac{jm}{r} H_m^{(2)}(kr)\hat{\theta} \right] J_m(k\rho) e^{jm(\phi-\theta)} \tag{3.67}$$

Since $\hat{\mathbf{r}}$ and $\hat{\theta}$ are orthogonal, the quantity $\frac{\partial g}{\partial \mathbf{n}}$ is obtained by taking the dot product of (3.66) and (3.67). The null-field equations for the Neumann case can then be derived in effectively the same way as for the Dirichlet case:

$$\begin{aligned} f(\theta) &= \psi(\sigma(\theta), \theta, k)/\sigma(\theta), \\ H_{mn} &= -C_m \int_0^{2\pi} \left[k\sigma^2(\theta) H_m^{(2)\prime}(k\sigma(\theta)) + jm\sigma'(\theta) H_m^{(2)}(k\sigma(\theta)) \right] \\ &\quad e^{(j(n-m)\theta)}\, d\theta, \\ J_{mn} &= -\int_0^{2\pi} \left[k\sigma^2(\theta) J_m'(k\sigma(\theta)) + jm\sigma'(\theta) J_m(k\sigma(\theta)) \right] \\ &\quad e^{(j(n-m)\theta)}\, d\theta. \end{aligned} \tag{3.68}$$

The null-field method has become widely accepted as a numerically efficient means of calculating the field from impenetrable and homogeneous objects of arbitrary shape (Bates and Wall, 1977; Varadan *et al.*, 1978; Ström and Zheng, 1987; Ström and Zheng, 1988).

3.4 Surface Integral Equation Method

It is pointed out by Colton (1984) that the technique described in this section, called the surface integral equation method, is closely related to the null-field method. It provides an exact integral representation of the field scattered by a surface on which either Dirichlet or Neumann boundary conditions are satisfied (Colton and Kress, 1983).

When Dirichlet boundary conditions are satisfied on $\sigma(\mathbf{x})$, the total field $\psi(\mathbf{x}, k)$ can be expressed, through (3.47), as a surface integral over equivalent sources on $\sigma(\mathbf{x})$:

$$\psi(\mathbf{x}, k) = -\int_\sigma \frac{\partial \psi(\mathbf{x}_1, k)}{\partial \mathbf{n}} g(\mathbf{x}, \mathbf{x}_1, k) d\sigma(\mathbf{x}_1) \tag{3.69}$$

where $\frac{\partial \psi(\mathbf{x}_1, k)}{\partial \mathbf{n}}$ is the equivalent source density. Consequently, because $\psi(\mathbf{x}, k) = 0$ on $\sigma(\mathbf{x})$,

$$\psi_i(\mathbf{x}, k) = \int_\sigma \frac{\partial \psi(\mathbf{x}_1, k)}{\partial \mathbf{n}} g(\mathbf{x}, \mathbf{x}_1, k) d\sigma(\mathbf{x}_1) \quad \text{for} \quad \mathbf{x} \in \sigma(\mathbf{x}) \tag{3.70}$$

which can be regarded as a Fredholm integral equation of the first kind for the equivalent source density.

A similar integral equation is found for the density of equivalent surface sources giving rise to a total field that satisfies Neumann boundary conditions on $\sigma(\mathbf{x})$:

$$\frac{\partial \psi_i(\mathbf{x}, k)}{\partial \mathbf{n}} = -\int_\sigma \psi(\mathbf{x}_1, k) \frac{\partial^2 g(\mathbf{x}, \mathbf{x}_1, k)}{\partial^2 \mathbf{n}} d\sigma(\mathbf{x}_1) \quad \text{for} \quad \mathbf{x} \in \sigma(\mathbf{x}) \tag{3.71}$$

In principle, both (3.70) and (3.71) may be solved by the algebraic expansion technique discussed in §2.2.8. They tend to be ill-conditioned, however, because they are integral equations of the first kind. Ill-conditioning also arises because their solutions are non-unique when k is an interior eigenvalue of the surface $\sigma(\mathbf{x})$. This has motivated modifications to (3.70) and (3.71) which allow them to be transformed into integral equations of the second kind and force their solutions to be unique for all k (Jones, 1974; Kleinman and Roach, 1974; Colton and Kress, 1983; Colton, 1984). Numerical results illustrating these modifications have been presented by Roger (1981) and Colton (1984).

3.5 Eigenfunction Expansion

Eigenfunction expansions provide exact analytic solutions of the Helmholtz equation. Such solutions are obtained for an object which can be partitioned into homogeneous regions whose boundaries correspond to surfaces or lines on which coordinates, belonging to any of the eleven separable systems (Morse and Feshbach, 1953, Chapter 5), are constant.

The foundation of the method is to find solutions of the Helmholtz equation in each homogeneous region of the form

$$\psi(\mathbf{x}, k) = F_1(x_1, \nu)F_2(x_2, \nu)F_3(x_3, \nu) \tag{3.72}$$

where x_1, x_2, x_3 represent a general right handed, curvilinear coordinate system and ν is the constant refractive index of the homogeneous region. Each of the terms in the expansion is known as a mode or partial wave. A solution of the form (3.72) is known as separable because it comprises eigenfunctions each dependent on only one coordinate. In any of the eleven separable coordinate systems, the Helmholtz equation has an infinite number of individual solutions of the form of (3.72). The most general solution in a homogeneous region is then a linear combination of the individual solutions. Since the three factors on RHS (3.72) are together often referred to as an eigenfunction, the expression

$$\psi(\mathbf{x}, k) = \sum_{m,n,p=-\infty}^{\infty} A_{mnp}F_{1,m}(x_1, \nu)F_{2,n}(x_2, \nu)F_{3,p}(x_3, \nu) \tag{3.73}$$

is called an eigenfunction expansion. Note that the A_{mnp}, which are constant expansion coefficients, are complex in general. Expansions in other regions of the object have the same general form. The eigenfunctions are orthogonal (Morse and Feshbach, 1953, Chapter 5), so that

$$\int_L F_{1,m}(x_1)F_{1,n}(x_1)dx_1 = C_n^{F_1}\delta_{mn} \tag{3.74}$$

where L represents the interval of the coordinate axis on which the eigenfunction $F_{1,m}(x_1)$ is defined and $C_n^{F_1}$ is a constant.

To find a solution to a particular problem, supplementary conditions must be invoked (see §2.2.5), in particular the boundary, radiation and finiteness conditions. The boundary conditions are satisfied by matching the general solutions from adjacent regions across their boundary surfaces, provided each surface spans the full range of the coordinates that vary along that surface. As an example, consider the regions Υ_+ and

Υ_- lying on each side of the surface $x_1 = a$. The boundary conditions are satisfied by

$$\begin{aligned} A^+_{mnp} F_{1,m}(a,\nu^+) &= A^-_{mnp} F_{1,m}(a,\nu^-) \\ \frac{\partial}{\partial x_1} A^+_{mnp} F_{1,m}(a,\nu^+) &= \frac{\partial}{\partial x_1} A^-_{mnp} F_{1,m}(a,\nu^-) \end{aligned} \tag{3.75}$$

for each mode m, n, p where the superscripts + and - denote the general solution in the regions Υ_+ and Υ_- respectively. Eigenfunctions of the Helmholtz equation can also exhibit infinities at points of the interval on which they are defined (Morse and Feshbach, 1953, Chapter 5). The finiteness condition can only be satisfied if the eigenfunctions of a solution are finite in the region in which the solution is defined. This is achieved by invoking only those eigenfunctions which are finite in that region. The radiation condition is satisfied similarly by choosing those eigenfunctions which are outgoing in the region in which the radiation condition must be satisfied.

There is an addition theorem for each of the aforesaid eigenfunctions (Morse and Feshbach, 1953, Chapter 5). These theorems allow translation of the coordinate origin, thereby increasing the number of objects for which exact analytic solutions can be obtained. The effect of applying the addition theorem for the cylindrical coordinate system is illustrated in fig. 3.10. The origin O of the coordinates $(r;\theta)$ is translated by $(R;\beta)$ to the origin O' of the coordinates $(\rho;\phi)$. The eigenfunctions in each of the coordinate systems for $R > \rho$ are related by (Watson, 1966, Chapter 11)

$$\mathcal{C}_m(r)e^{jm\theta} = \sum_{n=-\infty}^{\infty} \mathcal{C}_{m-n}(R) J_n(\rho) e^{jn\phi} e^{j(m-n)\beta} \tag{3.76}$$

where m and n are integers and $\mathcal{C}_m(\cdot)$ denotes any cylindrical function of order m. Note that the coordinates $(\rho;\phi)$ and $(R;\beta)$ must be interchanged on RHS (3.76) when $R < \rho$.

Because the eigenfunction expansion method allows exact analytic expansions for the field, it can be used as a reference to check the accuracy of the ray tracing and Rayleigh-Gans (Born) methods. To illustrate this, it is shown below how to invoke the eigenfunction expansion method to calculate the scattering from the standard object defined in fig. 3.3. Because the surface $\sigma(\mathbf{x})$ is a circle, eigenfunctions appropriate to cylindrical coordinates are invoked. $\psi_i(\rho;\phi)$ can then be expressed as (see §2.2.1)

$$\psi_i(\rho;\phi) = \sum_{m=-\infty}^{\infty} a_m \, J_m(k\rho) \, e^{jm\phi} \tag{3.77}$$

where the expansion coefficients a_m are given by (2.11). The scattered field in the region Υ_+ can be written as

$$\psi_s(\rho;\phi) = \sum_{m=-\infty}^{\infty} A^+_m \, H^{(2)}_m(k\rho) \, e^{jm\phi} \tag{3.78}$$

where the A^+_m are expansion coefficients. Implicit in (3.77) and (3.78) are the respective understandings that $\psi_i(\mathbf{x},k)$ must be finite and well behaved as $r \to 0$ (in cases for which the scattering region is infinitesimal, for instance) and that $\psi_s(\mathbf{x},k)$ is entirely outgoing in Υ_+.

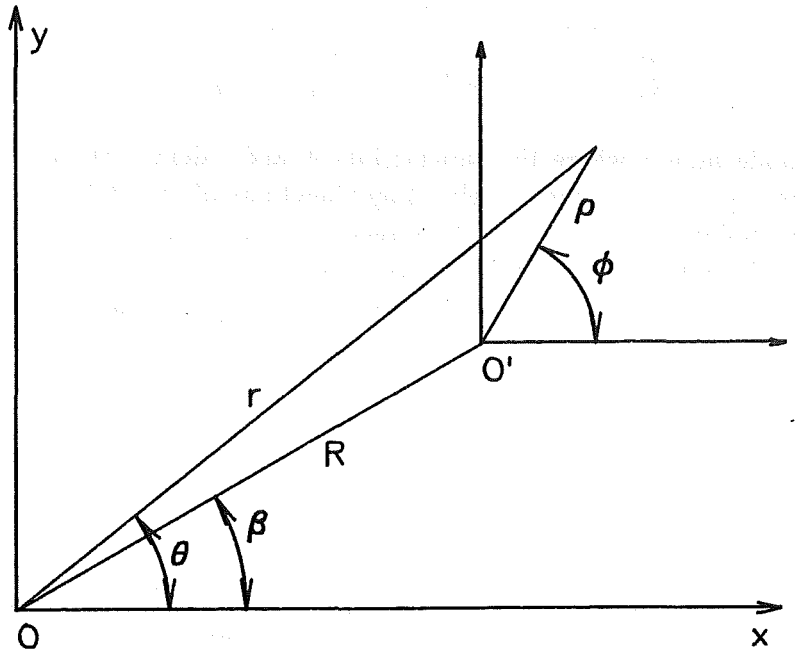


Figure 3.10: Translation of origins by the addition theorem in cylindrical coordinates. The origin O of the coordinates $(r;\theta)$ is translated by $(R;\beta)$ to the origin O' of the coordinates $(\rho;\phi)$.

For points in Υ_-, the allowable expansion for the field is

$$\psi(\rho;\phi) = \sum_{m=-\infty}^{\infty} A_m^- J_m(k\nu^-\rho)\, e^{jm\phi}, \tag{3.79}$$

because only the Bessel functions of the first kind are finite throughout Υ_-. Note that the A_m^- are constant expansion coefficients in the region Υ_-.

Finally the boundary conditions (3.75) are invoked to give

$$\begin{aligned} A_m^- J_m(k\nu^- R) &= A_m^+ H_m^{(2)}(kR) + (-j)^m J_m(KR) \\ \nu^- A_m^- J_m'(k\nu_- R) &= A_m^+ H_m'^{(2)}(kR) + (-j)^m J_m'(kR) \end{aligned} \tag{3.80}$$

These two equations allow each of the constants $A_m^{\pm}$ to be determined for all m. Hence a solution for the field can be obtained.

Fig. 3.11 and fig. 3.12 illustrate the application of the eigenfunction expansion method to the calculation of the field scattered from the standard object (see fig. 3.3). When these results are compared with the results obtained from ray-tracing (fig. 3.4) and the Born approximation (fig. 3.9) an indication of the applicability of these two approximations can be obtained. Ray-tracing is shown to be more effective when the diameter of the standard object is greater than at least 10 λ and also when the refractive index deviates from unity by less than 20%. The Born approximation becomes more effective when $2r_o|\nu_- - 1| < \frac{\lambda}{2}$. Although the eigenfunction expansion method provides exact analytic expressions for the field, its use is restricted to objects conforming to specialised geometries.

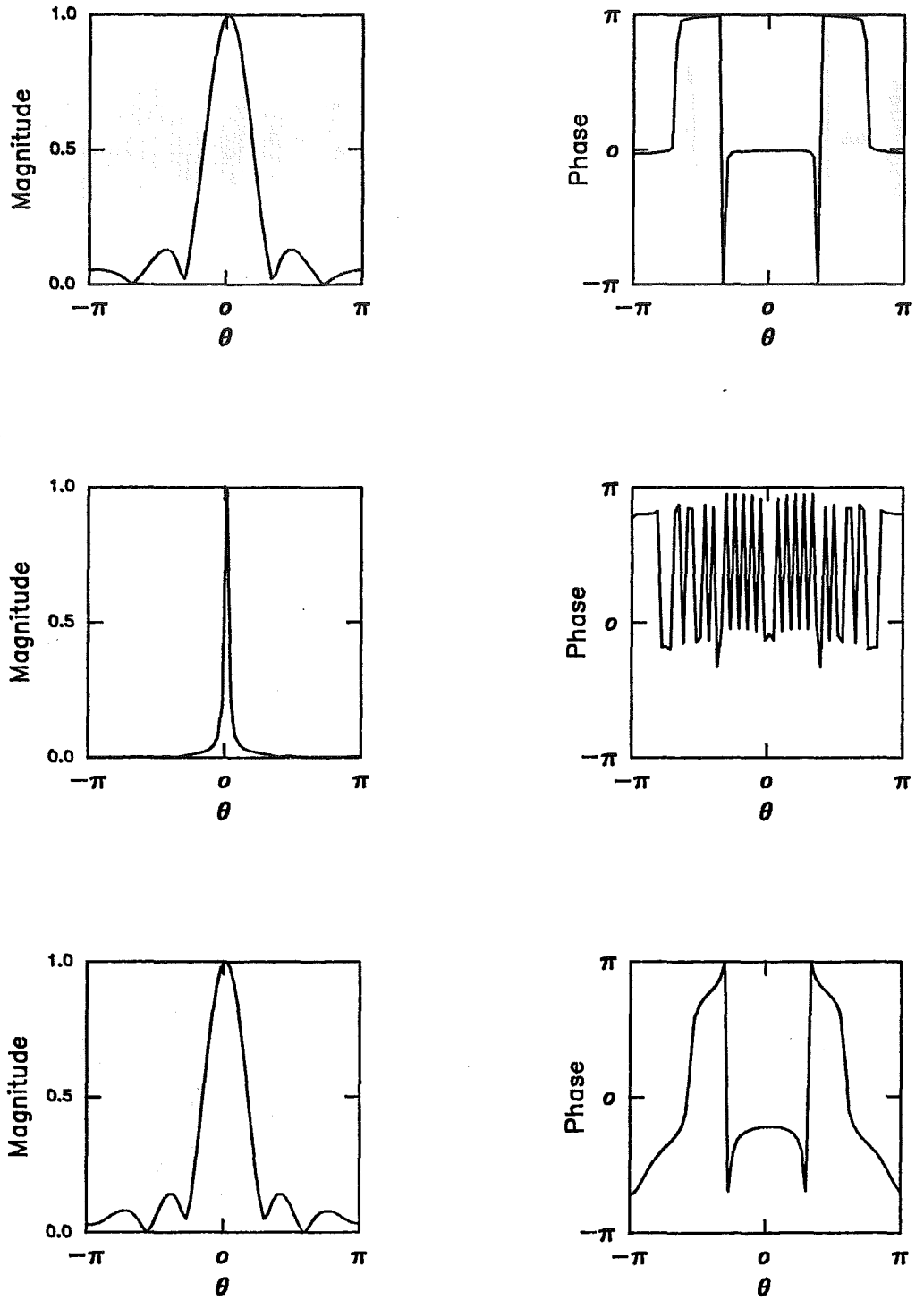


Figure 3.11: Illustration of application of the eigenfunction expansion method. The magnitude and phase of the field scattered from the standard object (shown in fig 3.3) are calculated using (3.80). The top, middle and bottom rows of graphs in the figure correspond, respectively, to the scattering object with $\nu_- = 1.01$ with $2r_o = 1.25\lambda$, $\nu_- = 1.01$ with $2r_o = 10\lambda$ and $\nu_- = 1.2$ with $2r_o = 1.25\lambda$.

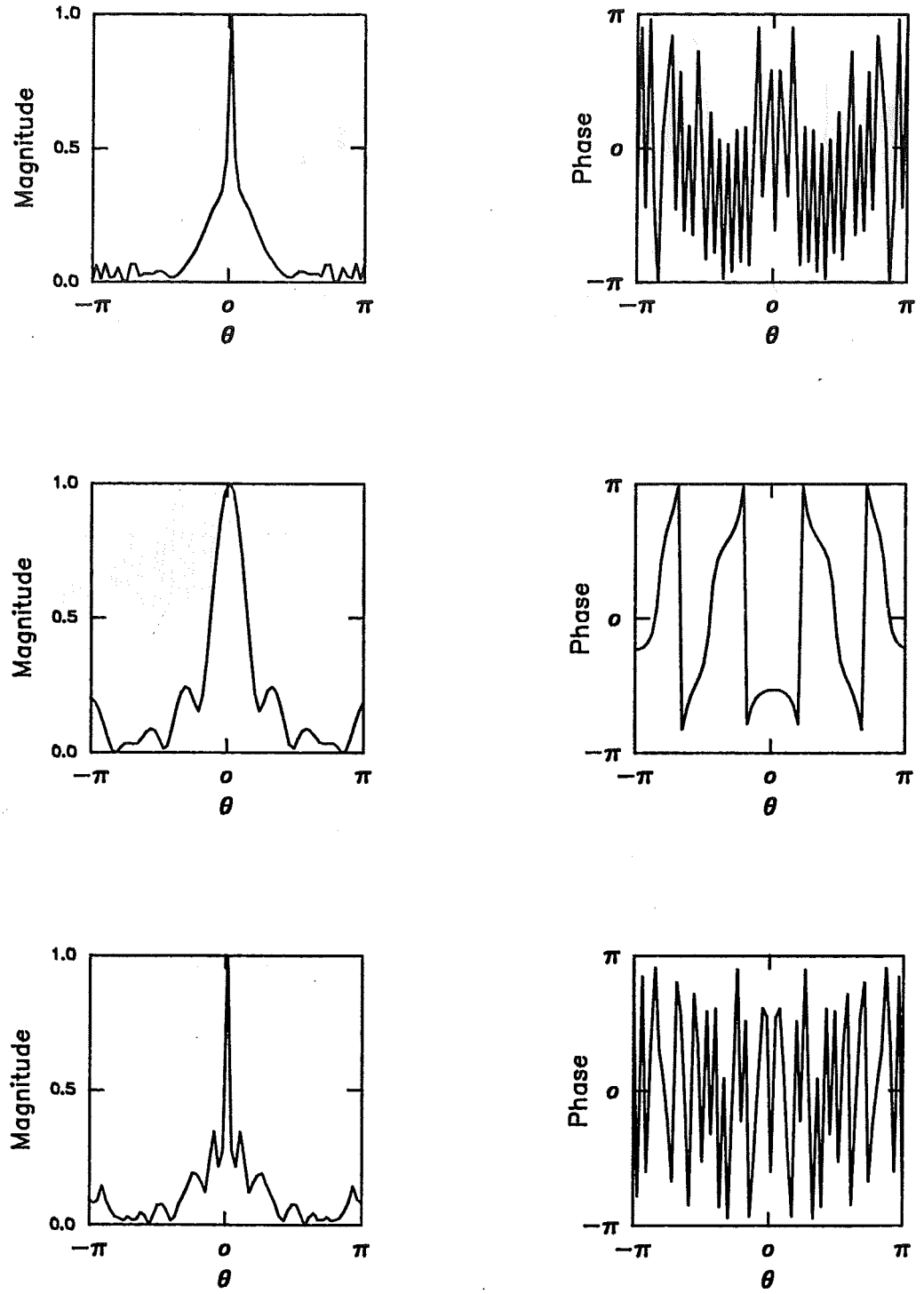


Figure 3.12: Further illustration of the application of the eigenfunction expansion method. The magnitude and phase of the field scattered from the standard object (shown in fig 3.3) are calculated using (3.80). The top, middle and bottom rows of graphs in the figure correspond, respectively, to the scattering object with $\nu_- = 1.2$ with $2r_o = 10\lambda$, $\nu_- = 1.5$ with $2r_o = 1.25\lambda$ and $\nu_- = 1.5$ with $2r_o = 10\lambda$.

3.6 The On-surface Radiation Condition Method

The on-surface radiation condition (OSRC) was recently developed by Kriegsmann *et al.* (1987) from the results of research into finite-difference time-domain methods for solving the direct problem (Kriegsmann *et al.*, 1987). OSRC allows the approximate solution of scattering from impenetrable and penetrable homogeneous objects. The method is most appropriate for objects whose linear dimensions, in wavelengths, are large enough that numerical solutions obtained via numerical evaluation of integral equations involve unacceptably heavy computational loads. (Jones, 1988b, 1988a) has recently generalised the technique to three dimensions and has also devised general formulas for the on-surface radiation condition.

When the field scattered from an object is expressed as an integral equation over the surface field, (3.45) becomes

$$\psi_s(\mathbf{x},k) = \int_\sigma \left[\psi_s(\mathbf{x}_1,k)\frac{\partial g(\mathbf{x},\mathbf{x}_1,k\nu)}{\partial \mathbf{n}} - g(\mathbf{x},\mathbf{x}_1,k\nu)\frac{\partial \psi_s(\mathbf{x}_1,k)}{\partial \mathbf{n}}\right] d\sigma(\mathbf{x}_1) \qquad (3.81)$$

The ORSC technique enables (3.81) to be converted into either a surface integral over known quantities or a simple ordinary differential equation. The technique is predicated on certain approximate expressions for both $\psi(\mathbf{x},k)$ and $\frac{\partial\psi(\mathbf{x},k)}{\partial \mathbf{n}}$ on the surface of the scatterer. These expressions are obtained by applying the radiation condition (see §2.2.5), which states that it is necessary for outgoing scattered waves to be extinguished without undergoing significant reflections, on the surface of the scatterer (Kriegsmann *et al.*, 1987). It is known that a convenient way to ensure that the radiation condition is approximately fulfilled in finite regions is to employ differential operators (Bayliss and Turkel, 1980; Kriegsmann and Morawetz, 1980). In the ORSC method these operators are applied directly on the surface of the scatterer. It is then hoped that reasonable approximate forms of $\psi(\mathbf{x},k)$ and $\frac{\partial\psi(\mathbf{x},k)}{\partial \mathbf{n}}$ can be found on $\sigma(\theta)$.

Consider, as an example, scattering from a totally reflecting circular cylinder of radius r_o, on whose surface the Dirichlet boundary condition is invoked. A radiation condition operator which can be applied to the surface of the circular cylinder is (Kriegsmann *et al.*, 1987)

$$B = \frac{\partial}{\partial r} + \frac{1}{2\rho} - jk \qquad (3.82)$$

where an arbitrary point in space is accorded the coordinates $(\rho;\phi)$. On $\sigma(\mathbf{x})$, the operator B gives rise to a field that extinguishes outgoing scattered waves without causing non-physical reflections. Consequently,

$$B\psi_s(r_o;\phi,k) = 0 \qquad (3.83)$$

which can be manipulated into

$$\frac{\partial\psi_s(r_o;\phi,k)}{\partial \mathbf{n}} = (jk - \frac{1}{2\rho})\psi_s(r_o;\phi,k) \qquad (3.84)$$

Since the Dirichlet condition is specified on $\sigma(\mathbf{x})$, $\psi_s(r_o;\phi,k) = -\psi_i(r_o;\phi,k)$

$$\frac{\partial\psi(r_o;\phi,k)}{\partial \mathbf{n}} = (\frac{1}{2\rho} - jk)\psi_i(r_o;\phi,k) \qquad (3.85)$$

which, when substituted into the surface integral in (3.81), allows the scattered field to be calculated.

Kriegsmann *et al.* (1987) have applied the ORSC method to circular and rectangular cylinders and to a strip of infinite height, finite width and zero thickness. They have compared the results to exact and physical optics solutions to show that ORSC produces a better approximation to the exact scattered field than physical optics does for these examples. Jones (1988b) has calculated the scattering from a sphere and compared his results to exact and geometrical optics solutions. He shows that ORSC is more effective than geometrical optics for sphere diameters of 1.5 and 10 wavelengths respectively.

Chapter 4

Review of Solutions to Inverse Scattering Problems

This Chapter reviews inverse methods involving linear wave motion. The review concentrates on methods which may lead to algorithms of practical use and which have been numerically simulated.

Inverse methods are invoked for estimating the constitutive parameters of a body, given scattering data resulting from emanations interacting with the body. The equations describing the interactions of the emanations with the body, which are introduced in Chapter 2, can be generalised as

$$\chi(\mathbf{x}) = \Lambda^{-1}\left(\psi_s(\mathbf{x}, k), \psi_i(\mathbf{x}, k)\right) \tag{4.1}$$

where Λ^{-1} is an operator mapping the scattering data $\psi_s(\mathbf{x}, k)$ to the generalised constitutive parameter $\chi(\mathbf{x})$. Inverse methods provide means of solving operator equations of the type (4.1) for $\chi(\mathbf{x})$. However, there are major conceptual and computational difficulties associated with solving (4.1). The operator Λ^{-1} is usually ill-posed. That is, solutions deduced from Λ^{-1} can be expected to be either non-unique, or non-existent, or to depend discontinuously on the data. Also, as shown in Chapter 3, Λ^{-1} is in general a nonlinear implicit mapping. Consequently, it is almost always impossible to devise analytical solutions to the inverse problem. Thus, iterative approaches or approximations have to be invoked.

A discussion of the uniqueness of the inverse problem is given in §4.1, wherein it is shown that the dimensionality of the scattering data is necessarily less than that of the quantities that one wishes to recover. This notion is invoked to indicate that the solution of the inverse source problem must in general be non-unique. It is also argued that, because solutions to inverse scattering problems must also satisfy the wave equation in Υ_-, such problems can in principle have unique solutions.

In some situations, approximate representations of the total field can be invoked to eliminate many of the difficulties associated with solving the inverse problem. These approximations often allow explicit solutions to the inverse problem to be found. §4.2 discusses explicit solutions to the inverse problem. In §4.2.1 an inverse method, in which the total field is approximately represented in terms of straight rays, is presented. Because, in many instances, this closely approximates reality, it can be considered an explicit exact method. §4.2.2 reviews explicit solutions, herein called explicit approximate solutions, which rely on approximations to the total field that are less realistic

(they can nevertheless be useful). Techniques for generating implicit solutions based on approximations, of varying degrees of realism, are reviewed in §4.3.4.

It is sometimes impossible to find any viable approach to solving the inverse problem explicitly. This may be because appropriate approximations are too elusive or those approximations which do suggest themselves are so imprecise as to be useless. More accurate descriptions of the total field must then be invoked. Because these descriptions are necessarily more intricate only implicit solutions of the inverse problem can then be obtained. §4.3.3 presents solutions to inverse problems which incorporate exact descriptions of the total field (note also, as indicated in the final sentence of the previous paragraph that approximate implicit solutions are also covered in this section). It is important to appreciate, however, that ill-posedness of the problem and the solution of nonlinear operator equations must be properly handled. §4.3.1 explains how the problem of ill-posedness can be overcome. The solution of nonlinear equations is discussed in §4.3.2. Exact implicit solutions are then reviewed in §4.3.3. The computational cost, however, of exact implicit solutions generally prohibits their practical use.

4.1 Uniqueness and the Dimensionality Difficulty

Appropriately interpreted, the volume source formulation, presented in §3.2, provides a conceptual understanding of the uniqueness of solutions to the inverse scattering problem. This formulation expresses the scattered field as

$$\psi_s(\mathbf{x},k) = \int_{\Upsilon_-} (\chi^2(\mathbf{x}_1,k)-1)\psi(\mathbf{x}_1,k)g(\mathbf{x},\mathbf{x}_1,k)\,d\Upsilon(\mathbf{x}_1) \tag{4.2}$$

which relates the equivalent sources, of density $(\chi(\mathbf{x})^2-1)\psi(\mathbf{x},k)$, to the scattered field $\psi_s(\mathbf{x},k)$. A difficulty in inverse scattering contexts is that the scattered field can only be observed in Υ_{++} whereas the equivalent sources lie inside Υ_-.

Before the inverse scattering problem can be solved, scattering data must be obtained. All the recoverable (within Υ_{++}) information about the equivalent sources can be obtained by observing $\psi_s(\mathbf{x},k)$ on a surface surrounding Υ_- (Bates, 1980). The reason for this can be demonstrated straightforwardly in two dimensions. The field $\psi_s(\mathbf{x},k)$ can then be expressed in an eigenfunction expansion appropriate for cylindrical coordinates as (see §3.5)

$$\psi_s(\mathbf{x},k) = \sum_{m=-\infty}^{\infty} A_m H_m^{(2)}(kr)e^{jm\theta} \tag{4.3}$$

If $\psi_s(\mathbf{x},k)$ is observed on a circle of radius R, the A_m can be found by invoking orthogonality of the exponential functions, i.e.

$$\int_0^{2\pi} \psi_s(kR;\theta)e^{-jm\theta}d\theta = H_m^{(2)}(kR)A_m \tag{4.4}$$

The scattering data are thus effectively one-dimensional. Scattering data obtained on more than one circle therefore provides no essential extra information, although they can of course ameliorate the effects of noise.

Consider an object which is comprised of circularly symmetric equivalent sources $\tilde{\sigma} = \tilde{\sigma}(r)$. Because the field radiating from this object must also be circularly symmetric, all the A_m must be zero except for A_0. By substituting (3.45) and (4.3) into (4.2)

and invoking (4.4) gives (Devaney and Sherman, 1982)

$$\int_0^R \tilde{\sigma}(r) J_0^{(2)}(kr) dr = A_0 \tag{4.5}$$

Clearly, knowledge of A_0 alone is insufficient to determine $\tilde{\sigma}(r)$ and consequently solutions to the problem are non-unique. The essential cause of the non-uniqueness is that $\psi_s(\mathbf{x}, k)$ is one-dimensional whereas $\tilde{\sigma}(r)$ is in general two-dimensional.

The argument given in the previous paragraph can of course be generalised to any number of dimensions in any of the separable coordinate systems by invoking (3.73). Consequently, the scattering data are always of dimension one less than the quantities one wishes to recover. This discrepancy is invoked by Bates (1984) to illuminate the intricies associated with solving the inverse scattering problem. Bates (1984) refers to the discrepancy as the dimensionality difficulty.

In the context of an inverse source problem the dimensionality difficulty necessarily indicates that its solution is non-unique. For inverse scattering problems, however, there is the extra constraint that the wave equation must be satisfied in Υ_-. This suggests that unique solutions of the inverse scattering problem are possible (Devaney and Sherman, 1982). Because the wave equation must be satisfied in Υ_- the total field may in principle be continued back into Υ_- using the scattering data observed in Υ_+ (Bates, 1975). The only unknown then is the constitutive parameter, which is invariant with changes in either the direction or frequency of the incident field. For each incident field, however, the scattering data are in general different. This indicates that it is possible to restore a dimension to the inverse scattering problem by varying the incident field. Thus solutions to the inverse scattering problem can, provided any other necessary conditions are satisfied, be unique.

4.2 Explicit Solutions

When they can be usefully invoked, explicit solutions to the inverse scattering problem are generally computationally efficient. Consequently, these solutions tend to be useful in practice. Explicit solutions are usually obtained by employing an approximate representation of the total field. In some circumstances, however, the approximation may describe reality very closely, so that its invocation can lead to what can be considered an explicit exact solution.

4.2.1 Review of Explicit Exact Solutions

4.2.1.1 Computed Tomography

Consider a field describable by straight rays transversing an inhomogeneous medium embedded in Υ_- characterised by a constitutive parameter $\chi(\mathbf{x})$. The purpose of computed tomography (CT) is to reconstruct $\chi(\mathbf{x})$ from recorded values of the field after it has passed through Υ_- (Bates *et al.*, 1990). Although scattering of the field is appreciable in general, the interest lies in the field that passes undeflected through Υ_- provided it accommodates the properties of $\chi(\mathbf{x})$ through variable delays and/or attenuations (or whatever) due to $\chi(\mathbf{x})$. CT, which manifests itself in many technical sciences (Herman, 1979; Herman, 1983; Robb, 1985; Daily, 1986), is based on the scattering data

being representable as line integrals of $\chi(\mathbf{x})$ along straight rays traversing Υ_-, from Υ_{inc} (e.g. source or transmitter) to Υ_{obs} (e.g. sink or detector) of the field.

Fig. 4.1 shows two systems (x, y, z) and (ξ, η, ζ), of Cartesian coordinates having the common origin O. The systems can be thought of as coinciding initially with the z-axis then being rotated by θ, keeping the z, x-plane and the ζ, ξ-plane coincident. The dashed line joining the source and sink in fig. 4.1 is parallel to the ζ-axis. Each piece of CT data is thought of as the line integral of $\chi(\mathbf{x})$ along such a dotted line. The two-dimensional projection $p(\xi, \eta; \theta)$ at angle θ is defined by

$$p(\xi, \eta; \theta) = \int_{-\infty}^{\infty} \chi(\mathbf{x})\, d\zeta \tag{4.6}$$

where the infinite limits merely emphasize that the straight ray, between source and sink (wherever they are positioned), passes right across Υ_-. On recalling the Fourier transform relationships defined in §2.2.6, the projection theorem in three dimensions is obtained (Barrett, 1981)

$$p(\xi, \eta; \theta) = \mathcal{F}_2^{-1}\left[\tilde{\chi}(\alpha, \beta; \theta)\right] \tag{4.7}$$

where $\tilde{\chi}$ is the three-dimensional Fourier transform of χ and the α, β-plane in Fourier space corresponds to the ξ, η-plane in image space. Since, for all θ from 0 to π, the ξ, η-plane spans all of three-dimensional space, if $p(\xi, \eta; \theta)$ is given for all θ, it is seen from (4.7) that the complete three-dimensional spectrum $\tilde{\chi}(\mathbf{u})$ of $\chi(\mathbf{x})$ can be recovered. So, $\chi(\mathbf{x})$ can in principle be reconstructed from a complete set of projections. Because the volume of Υ_- is finite, the Fourier sampling theorem (see §2.2.7) indicates that only a finite set of projections is needed, at a discrete set of values for the angle θ. The two-dimensional manifestation of the theory outlined above has received more practical attention (Herman, 1980; Bates and McDonnell, 1989, Chapter 5). This is because the rays, being straight, can be confined to a single plane by appropriate positioning of the source and sink. We relate two-dimensional versions of $\chi(\mathbf{x})$ and $\tilde{\chi}(\mathbf{u})$ through

$$\mathcal{F}_2\left[\chi(x, y)\right] = \tilde{\chi}(u, v) = F(\rho; \phi) \tag{4.8}$$

The one-dimensional projection at angle ϕ is defined, with the aid of the coordinates introduced in fig. 4.2, by

$$p(\xi; \phi) = \int_{-\infty}^{\infty} \chi(x, y)\, d\eta \tag{4.9}$$

and the two-dimensional projection theorem states that

$$p(\xi; \phi) = \mathcal{F}_1^{-1}\left[\tilde{\chi}(\alpha; \phi)\right] \tag{4.10}$$

where ξ and α are both taken to range from $-\infty$ to ∞, which indicates that the one-dimensional Fourier transform (or spectrum) of a projection represents the complex amplitude, along the straight line passing through the origin of Fourier space at the same angle to the u-axis as the ξ-axis is in image space to the x-axis, of the complete two-dimensional spectrum of the image. Because Υ_- is finite, $\tilde{\chi}(u, v)$ is fully characterised by a finite number of projections, which are best recorded at a set of equi-spaced angles (Herman, 1980, Chapter 9; Lewitt, 1983). It is worth emphasising that these projections define $\tilde{\chi}(u, v)$ on a radial grid in Fourier space. Fourier transformation of $\tilde{\chi}(u, v)$ then leads to a reconstruction of the constitutive parameter $\chi(\mathbf{x})$. Once sufficient samples of the Fourier transform of $\chi(\mathbf{x})$ have been obtained a means

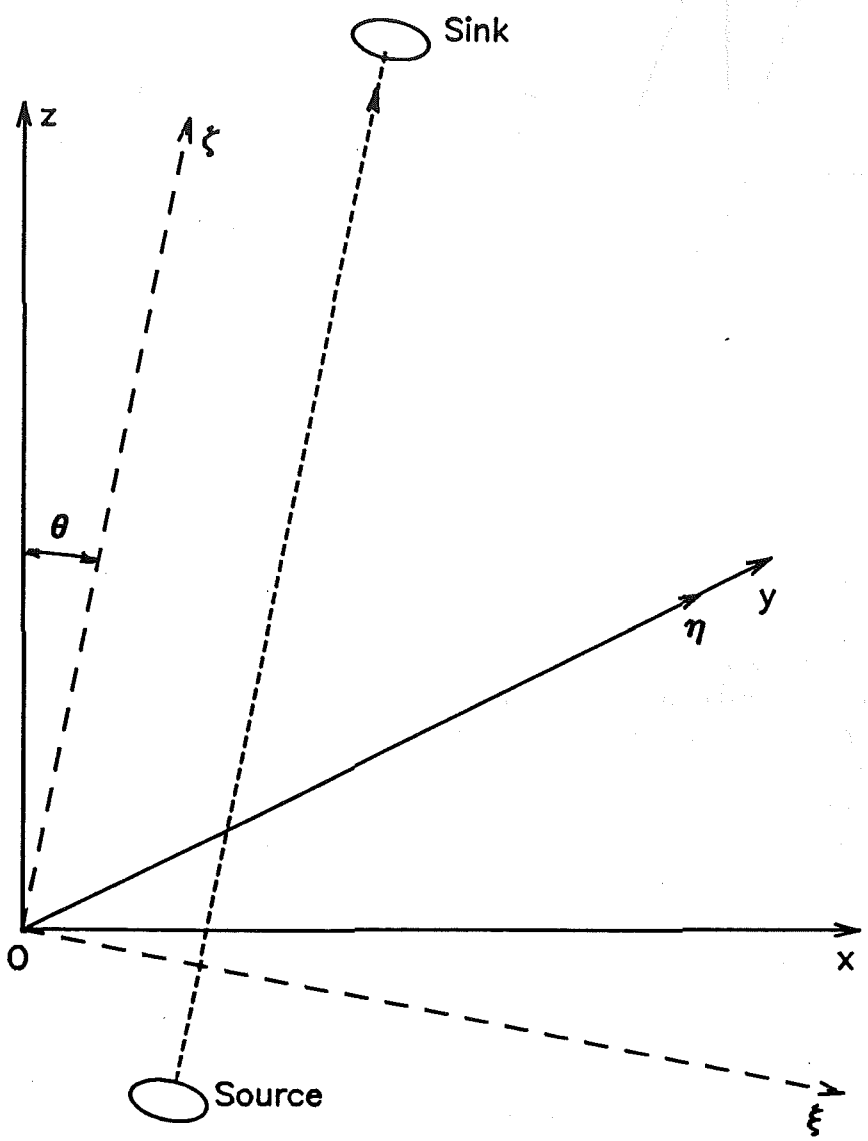


Figure 4.1: Coordinate systems and notational details for computed tomography in three dimensions. The dashed line is parallel to the ζ-axis, which lies in the same plane as the $\xi-, z-$ and x-axes.

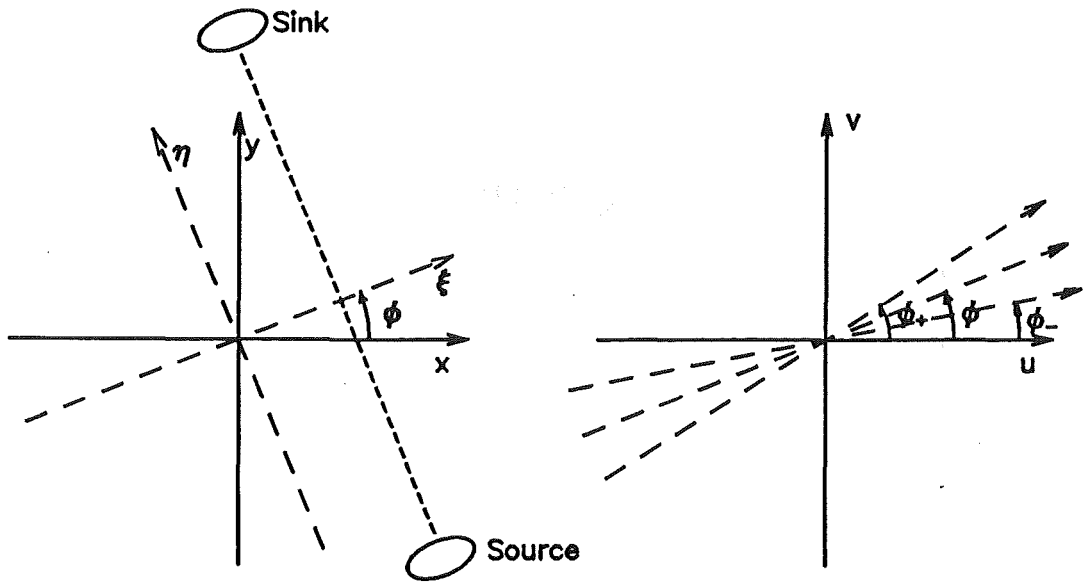


Figure 4.2: Coordinate systems and notational details for computed tomography in two dimensions. Image space and Fourier space are on the left and right respectively. The dashed line in image space passes across Υ_-. The dashed lies in Fourier space represent the boundaries of the double-sector discussed in §4.2.1.1.

of reconstructing $\chi(\mathbf{x})$ is required. Because of the high numerical efficiency of the FFT algorithm it is sensible to try to employ it, and when this is done it is refered to as the Fourier reconstruction procedure (Herman, 1979, Chapter 9). The FFT algorithm requires the data samples to be specified on a rectangular grid (Cooley and Tukey, 1965; Brigham, 1974). Because CT data constitute samples of $\tilde{\chi}(u,v)$ on a radial grid, the FFT algorithm cannot be invoked straightforwardly. Interpolation of the data from the radial grid onto a rectangular grid might seem an obvious solution to this difficulty. However, the resulting interpolation errors are generally unacceptable in practice. A method known as modified back-projection, which overcomes the interpolation problem and is now commonly used in practice (Lewitt, 1983; Bates and McDonnell, 1989, Chapter 5), as described below.

The three dashed lines in Fourier space, shown in fig. 4.2, correspond to the straight lines on which $\tilde{\chi}(u,v)$ is characterised by projections at angles ϕ_-, ϕ and ϕ_+ (note that these angles define three of the radial lines comprising the above-mentioned radial grid), where

$$\phi_+ - \phi = \phi - \phi_- = \Delta\phi \tag{4.11}$$

when there are $\pi/\Delta\phi$ equi-spaced projections. All of Fourier space can then be spanned with a finite number of straight lines by spreading each one out into a double sector of width $\Delta\phi$. For projections at the set $\{n\pi/N; n = 1, 2, \cdots, N\}$ of angles, the spectrum is approximated by

$$F(\rho;\phi) \approx (|\rho|/\rho) \sum_{n=1}^{2N} F(\rho; n\pi/N)\, \delta(\phi - n\pi/N) \tag{4.12}$$

where each double-sector has been separated into its two halves (e.g. for $\phi = n\pi/N$ and $\phi = (N+n)\pi/N$). Note that $\rho^{-1}\,\delta(\phi - n\pi/N)$ isolates one half of the n^{th} straight

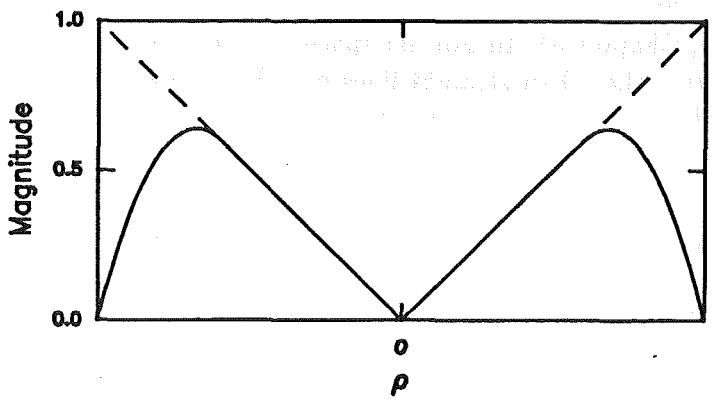


Figure 4.3: Windowed $|\rho|$ function. The dashed and solid lines represent the $|\rho|$ and windowed $|\rho|$ functions respectively.

line, while the factor $|\rho|$ effectively spreads it out into a sector. Taking the inverse two-dimensional Fourier transform of (4.12) gives (Herman, 1980, Chapter 9; Bates and McDonnell, 1989, Chapter 5)

$$\chi(x,y) = \sum_{n=1}^{N} \tilde{p}(\xi, n\pi/N) \tag{4.13}$$

for all points $(\xi,\eta) \in \Upsilon_{-}$, with the modified projection $\tilde{p}(\xi;\phi)$ defined by

$$\tilde{p}(\xi;\phi) = p(\xi;\phi) \odot h(\xi) \tag{4.14}$$

where $\odot$ is the one-dimensional convolution operator and $h(\xi)$ is the one-dimensional inverse Fourier transform of $|\rho|$. $h(\xi)$ is referred to as a convolutional filter. The convolution operation in (4.14) can also be performed by invoking the one-dimensional Fourier transform so that

$$\tilde{p}(\xi;\phi) = \mathcal{F}_1\left[\mathcal{F}_1^{-1}[p(\xi;\phi)]|\rho|\right] \tag{4.15}$$

In practice the frequency content of the image is only appreciable within a finite region. Because $|\rho|$ assumes correspondingly larger values at frequencies outside this region small errors in the data can be magnified considerably. To overcome this difficulty practical implementations of $|\rho|$ are smoothly rolled-off outside the region of Fourier space within which $|\tilde{\chi}(\alpha;\phi)|$ has sensible value. This roll-off is conveniently effected by a windowing function (Brigham, 1974), as illustrated in fig. 4.3. Each term in the summation in (4.13) represents what is called a back-projection, because the value of $\tilde{p}(\xi; n\pi/N)$ is stored at each pixel (in image space) intersected by the line parallel to the η-axis and distant ξ from it (refer to fig. 4.1 and take $\phi = n\pi/N$). The formula (4.14) is known as modified, or filtered (Lewitt, 1983), back-projection since each of the N original projections is modified in the manner intimated by (4.14).

The above discussion implicitly assumes that $p(\xi;\phi)$ is arbitrarily finely resolved, which is the same thing as saying that individual rays, each of infinitesimal width, of $\psi(\mathbf{x},k)$ can be separately observed. Resolution is always finite in practice, of course, implying that it is only possible to observe $p(\xi;\phi) \odot m(\xi)$, where $m(\xi)$ is some data function having a finite effective width, $\Delta\xi$ say (Herman, 1980, Chapter 3; Bates and McDonnell, 1989, Chapter 5). In Fourier space, the corresponding radial grid is composed of segments rather than straight lines of infinitesimal width. The resolution of the reconstructed image cannot be finer than $\Delta\xi$. In practice, however, the finiteness of N tends to limit the resolution more than the finiteness of $\Delta\xi$.

4.2.2 Review of Explicit Approximate Solutions

A variety of approximate approaches exist for obtaining an explicit solution to the inverse problem. Typically, they involve invoking an approximation to the field in Υ_-. The advantage of this approach is that problems associated with stability, the dimensionality difficulty and the solution of nonlinear equations are overcome.

4.2.2.1 Rayleigh-Gans (Born) Inversion

Rayleigh-Gans (Born) inversion explicitly inverts the first order Rayleigh-Gans (Born) approximate solution to the direct problem §3.2.1. Rayleigh-Gans (Born) inversion is also referred to as diffraction tomography (Devaney, 1982).

The first order Rayleigh-Gans (Born) approximation relates the scattered field $\psi_{ff_s}(\hat{\mathbf{x}},k)$ to the Fourier transform of the constitutive parameter $\chi(\mathbf{x})$ (3.32):

$$\psi_{ff_s}(\hat{\mathbf{x}},\mathbf{k}) = k^2\mathcal{F}_2\left[\chi^2(\mathbf{x})-1\right](\mathbf{k}') \tag{4.16}$$

where $\mathbf{k}' = k(\hat{\mathbf{k}}-\hat{\mathbf{x}})$.

Inspection of (4.16) might suggest that $\chi(\mathbf{x})$ could be immediately and straightforwardly recovered from the scattered field by invoking the inverse Fourier transform of $\psi_{ff_s}(\hat{\mathbf{x}},k)$. However, it is apparent from §3.2.1 and fig. 3.8 that $\psi_{ff_s}(\hat{\mathbf{x}},k)$, for all $\mathbf{x}$, provides information about $\mathcal{F}_2(\chi^2(\mathbf{x})-1)$ only on a circle. Consequently, in order to cover as much of Fourier space as possible, observations of $\psi_{ff_s}(\hat{\mathbf{x}},k)$ must be made for all $\hat{\mathbf{k}}$ as well. The coverage of Fourier space then corresponds to the interior of a circle Fourier space of radius $2k$ (see §3.2.1). Thus, the spatial resolution Δx of the image is

$$\Delta x \leq \frac{2\pi}{4k} = \frac{\lambda}{2} \tag{4.17}$$

which accords with the intuitive notion that spatial variations of the constitutive parameter significantly smaller than a wavelength cannot be discerned. The inverse Fourier transform of the scattering data is then a band limited approximation to the true distribution of $\chi(\mathbf{x})$.

The faithfulness of the reconstructed image is limited by the approximations invoked in the Rayleigh-Gans (Born) approximation. It is shown in §3.2.1 that the Rayleigh-Gans (Born) approximation can be expected to be valid if the refractive index satisfies $|\int_{l_1}^{l_2}\nu dl - L| < \lambda/2$ for all incident rays transversing Υ_-, where l_1 and l_2 identify the points where the incident ray enters and leaves Υ_-. Consequently, Rayleigh-Gans

(Born) inversion must produce distorted images for objects not satisfying the above condition. It should also be noted that the lack of explicit accommodation in the Rayleigh-Gans (Born) of multiple scattering and attenuation also introduce distortions. These effects have been investigated by Slaney *et al.* (1984), Soumekh and Kaveh (1986). Despite a lack of justification in many instances, the Rayleigh-Gans (Born) approximation remains one of the most commonly invoked approximations (Boyse and Keller, 1986; Blackledge *et al.*, 1987; Wu and Toksöz, 1987; Pratt and Worthington, 1988) due to the ease with which its inverse can be formed.

Numerical techniques for inverting (4.16) based on the FFT algorithm require that the scattering data be sampled on a rectangular grid §4.2.1.1. This can be realised by making observations of the field at specified $\hat{\mathbf{x}}$ and $\hat{\mathbf{k}}$ so that each sample lies at a point on the desired grid. However, for each grid point a different $\hat{\mathbf{x}}$ and $\hat{\mathbf{k}}$ must be employed, which places an unfortunate burden on the data gathering apparatus. Interpolation could be invoked to transfer the data on to the rectangular grid. However, unless performed accurately, and therefore protractedly, interpolation generates artefacts in the image (see §4.2.1.1). To overcome this problem, Devaney (1982) developed the back-propagation algorithm. It has many similarities to the back-projection algorithm and is discussed in §4.2.2.3.

Because Rayleigh-Gans (Born) inversion is one of the most commonly invoked inverse methods, an example of its application is now presented. Fig. 4.4 illustrates the application of (4.16) to the reconstruction of the standard object (defined in fig. 3.3). Cross-sections of reconstructions for two cases are depicted. The reconstructions are for $\nu_- = 1.01$ with $2r_o = 10\lambda$ and $\nu_- = 1.20$ with $2r_o = 10\lambda$. These results provide further illustration of the fact that the faithfulness of Rayleigh-Gans (Born) inversion decreases markedly when $2r_o|\nu_- - 1|$ exceeds $\frac{\lambda}{2}$.

4.2.2.2 Rytov Inversion

Rytov inversion is based on explicitly inverting the Rytov approximate solution to the direct problem which is

$$\tilde{\psi}(\hat{\mathbf{x}}, k) = k^2 \mathcal{F}_2 \left[\chi^2(\mathbf{x}) - 1\right](\mathbf{k}') \tag{4.18}$$

invoking the notation established in §3.2.2 and (3.43), where $\tilde{\psi}(\hat{\mathbf{x}}, k)$ is defined by (3.44) and $\mathbf{k}' = k(\hat{\mathbf{k}} - \hat{\mathbf{x}})$.

If (4.18) is compared with (4.16), it is clear that the same relationship exists between $\chi(\mathbf{x}, k)$ and $\tilde{\psi}(\hat{\mathbf{x}}, k)$ as for the Rayleigh-Gans (Born) approximation. Consequently, after first transforming $\psi_{ff_s}(\hat{\mathbf{x}}, k) \rightarrow \tilde{\psi}(\hat{\mathbf{x}}, k)$ by invoking (3.44), the same techniques can be applied to the Rytov inversion as to Rayleigh-Gans (Born) inversion.

The important differences between the Rytov and Rayleigh-Gans (Born) methods are their ranges of validity. As stated in §3.2.2, it is generally considered that the Rytov is more widely applicable.

4.2.2.3 The Back-propagation Method

Back-propagation is a numerically efficient method of performing Rayleigh-Gans (Born) and Rytov inversion. As mentioned in §4.2.2.1, it overcomes problems associated with interpolating the scattering data onto a rectangular grid. The method was originally

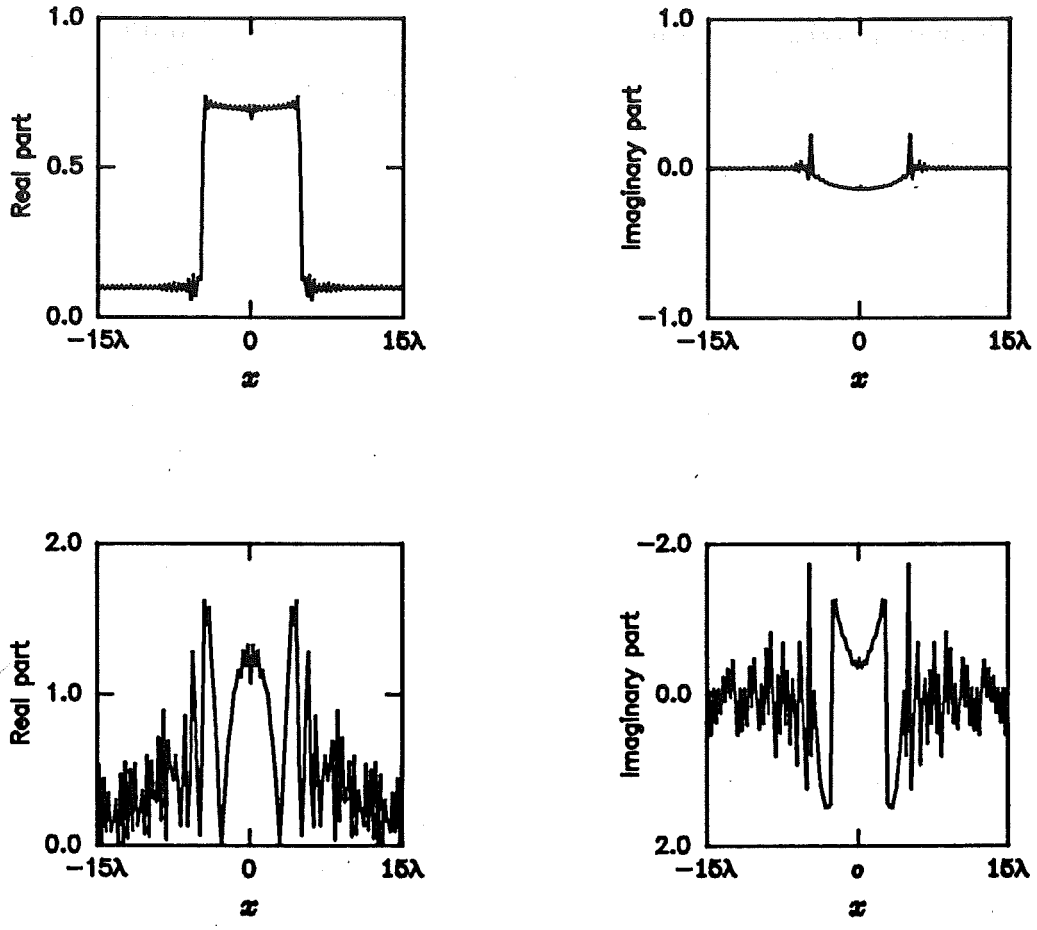


Figure 4.4: Illustration of the application of Rayleigh-Gans (Born) inversion. The top and bottom graphs in the figure correspond, respectively, to cross-sections along the x-axis of reconstructions of the standard object (defined in fig. 3.3) for which $\nu_- = 1.01$ with $2r_o = 10\lambda$ and $\nu_- = 1.20$ with $2r_o = 10\lambda$.

developed by Devaney (1982), who also has reviewed and improved it in a series of later papers (Devaney, 1983; Devaney, 1984; Devaney, 1985a; Devaney, 1985b; Devaney, 1986; Devaney, 1987; Devaney, 1989).

The back-propagation method is based on the Fourier transform, the formula (2.22) for which is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^K} \int (K) \int F(\mathbf{u}) e^{j\mathbf{u}\cdot\mathbf{x}} d\mathbf{u} \tag{4.19}$$

The quantity $f(\mathbf{x})$, which here represents $k^2(\chi^2(\mathbf{x}) - 1)$, is to be reconstructed from $F(\mathbf{u})$, which here represents scattering data $\psi_{ff_s}(\hat{\mathbf{x}}, k)$ for all $\hat{\mathbf{x}}$ and $\hat{\mathbf{k}}$. $F(\mathbf{u})$ is related to the scattering data $\psi_{ff_s}(\hat{\mathbf{x}}, k)$ through (see §3.2.1)

$$\psi_{ff_s}(\hat{\mathbf{x}}, k) = F\left(k(\hat{\mathbf{k}} - \hat{\mathbf{x}})\right) \tag{4.20}$$

As explained in §4.2.2.1 and §4.2.2.2, the FFT implementation of (4.19) for the Rayleigh-Gans (Born) or Rytov inversions cannot be invoked straightforwardly without causing unacceptable interpolation errors. In order to overcome this interpolation problem a change of variables is introduced:

$$\mathbf{u} = k(\hat{\mathbf{k}} - \hat{\mathbf{x}}) = \begin{pmatrix} k(\cos\phi_i - \cos\phi_s) \\ k(\sin\phi_i - \sin\phi_s) \end{pmatrix} \tag{4.21}$$

so that

$$d\mathbf{u} = k^2 \sqrt{1 - \cos^2(\phi_i - \phi_s)}\, d\phi_i d\phi_s \tag{4.22}$$

where the angles ϕ_i and ϕ_s represent the directions of $\hat{\mathbf{k}}$ and $\hat{\mathbf{x}}$ respectively as illustrated in fig. 4.5. The change of variables allows (4.19) to be written as

$$f(\mathbf{x}) = \frac{k^2}{2(2\pi)^2} \int_{-\pi}^{\pi} d\phi_i \int_{-\pi}^{\pi} d\phi_s \sqrt{1 - (\hat{\mathbf{k}} \cdot \hat{\mathbf{x}})}\, F(k(\hat{\mathbf{k}} - \hat{\mathbf{x}}))\, e^{jk(\hat{\mathbf{k}} - \hat{\mathbf{x}})} \tag{4.23}$$

which is seen to be ideally suited for Rayleigh-Gans (Born) inversion (and Rytov inversion after invoking (3.44)) (Devaney, 1982). For each ϕ_i, the integral over ϕ_s involves $F(\mathbf{u})$ evaluated on the locus of points implied in RHS (4.20). To conveniently express $f(\mathbf{u})$ in terms of $\psi_{ff_s}(\hat{\mathbf{x}}, k)$, the (κ, γ) coordinate system is introduced. This is related to the u, v axis by a rotation through $\phi_i + \frac{\pi}{2}$ as illustrated in fig. 4.5. Relative to (κ, γ)

$$\begin{aligned} \cos\left(\phi_s - (\phi_i + \frac{\pi}{2})\right) &= \kappa/k \\ \sin\left(\phi_s - (\phi_i + \frac{\pi}{2})\right) &= \frac{\sqrt{k^2 - \kappa^2}}{k} \end{aligned} \tag{4.24}$$

where $-k \leq \kappa \leq k$. For (4.24) to be valid, $\phi_s - \phi_i - \frac{\pi}{2}$ must lie in the interval $[0, \pi]$. Consequently, $F(\mathbf{u})$ must be bandlimited to lie within the circle of radius $< \sqrt{2}k$. It follows from (4.20) and (4.24) that

$$f(\mathbf{x}) == \frac{k}{2(2\pi)^2} \int_{-\pi}^{\pi} d\phi_i \int_{-\sqrt{2}k}^{\sqrt{2}k} \frac{d\kappa}{\gamma} |\kappa|\, \psi_{\phi_i}(\kappa, k)\, e^{jk(\kappa\xi + (\gamma - k)\eta)} \tag{4.25}$$

where ξ and η are defined in fig. 4.5 and $\psi_{\phi_i}(\kappa, k)$ denotes the scattered field $\psi_{ff_s}(\hat{\mathbf{x}}, k)$ due to an incident field in the ϕ_i direction, with κ and $\mathbf{x}$ being related by (4.21) and

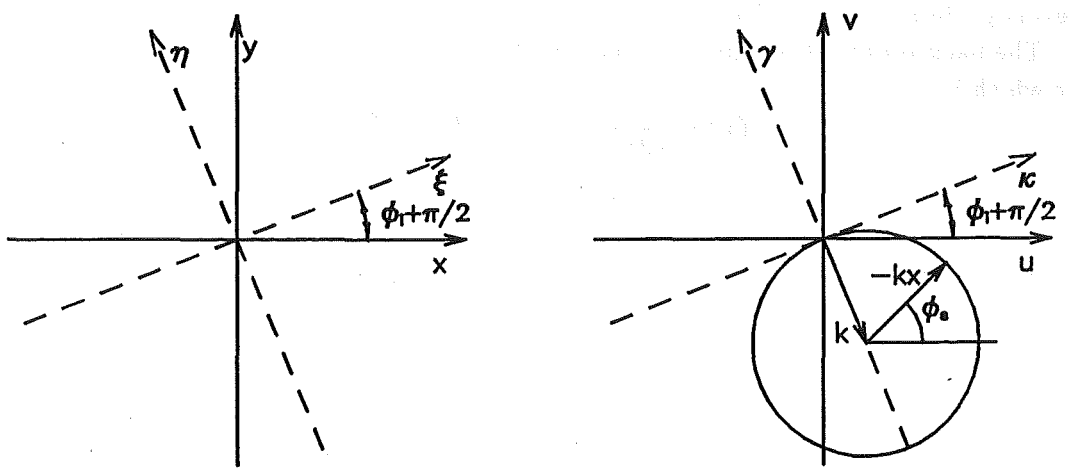


Figure 4.5: Coordinate systems and notational details for the back-propagation method. The x, y and u, v coordinate systems are in image and Fourier space respectively. The ξ, η and κ, γ coordinate systems are rotated by $\phi_i + \frac{\pi}{2}$ in image and Fourier space respectively.

(4.24). If $\psi_{\phi_i}(\kappa, k)$ is sampled uniformly with respect to κ then (4.25) can be implemented straightforwardly with the aid of the FFT, since interpolation is not required (Devaney, 1983). The quantities $|\kappa|/\gamma$ and $e^{jk(\kappa\xi+(\gamma-k)\eta)}$ are appropriately called filter and back-propagation operators respectively (Devaney, 1982).

The similarity of back-propagation to back-projection can be emphasised by writing

$$\Pi_{\phi_i}(\xi, \eta) = \int_{-\sqrt{2}k}^{\sqrt{2}k} \frac{d|\kappa|}{\gamma} |\kappa| \, \psi_{\phi_i}(\kappa, k) \, e^{jk(\kappa\xi+(\gamma-k)\eta)} \tag{4.26}$$

so that

$$f(\mathbf{x}) = \frac{k}{2(2\pi)^2} \int_{-\pi}^{\pi} d\phi_i \Pi_{\phi_i}(\xi, \eta) \tag{4.27}$$

which indicates that $\Pi_{\phi_i}(\xi, \eta)$ and $|\kappa|/\gamma$ can be likened to a modified projection and the convolutional filter respectively.

Devaney (1983; 1985b; 1987) has published several reconstructions illustrating his back-propagation technique. These reconstructions are excellent whenever errors involved in invoking the Rayleigh-Gans (Born) and Rytov approximations are small and they demonstrate that interpolation errors can be negligible. However, it must be remembered that the method can only perform as well as allowed by the Rayleigh-Gans (Born) and Rytov approximations. To round out this discussion, it is worth mentioning that an alternative back-propagation algorithm, which markedly improves computational efficiency with only a small sacrifice in overall image quality, has also been developed (Devaney, 1987).

4.2.2.4 The Causal Generalised Back-projection Method

The causal generalised back-projection method, due to Beylkin (1985a), allows reconstruction of discontinuous parts of a medium. It is thus particularly suitable for incorporation into techniques such as seismic migration imaging and medical B-scan (Beylkin, 1985b). Following Beylkin (1985a), other authors have extended and implemented the method (Schultz and Jaggard, 1987; Levy and Esmersoy, 1988).

Causal generalised back-projection is based on the distorted wave Rayleigh-Gans (Born) approximation (Taylor, 1972; Devaney and Oristaglio, 1983). This approximation is predicated on the constitutive parameter of the medium being decomposable into a known background and a perturbation. An example of such an object is the human breast. Its average refractive index is more or less known and one is interested in detecting deviations caused by, say, cancerous tumours (Mueller *et al.*, 1979). Knowledge of the background can then be used to approximate the total field in Υ_- by the scattering due to the background constitutive parameter. The perturbation can then be more accurately reconstructed than under the conventional Rayleigh-Gans approximation, for which the total field in Υ_- is approximated, within the integral describing the scattered field, by the incident field. To implement his method, Beylkin invokes scattering data observed over a band of frequencies and at different transmitter and receiver positions.

Assume that the constitutive parameter can be described as $\nu^2(\mathbf{x}) = \nu_0^2(\mathbf{x}) + \nu_1^2(\mathbf{x})$, where $\nu_0(\mathbf{x})$ and $\nu_1(\mathbf{x})$ are the known background and perturbed refractive index distributions respectively. By invoking the volume source formulation (3.28), the field scattered by $\nu_1^2(\mathbf{x})$ can be written as

$$\psi_s(\mathbf{x}, k) = \int_{\Upsilon_-} k^2(\nu_1^2(\mathbf{x}_1) - 1)\psi(\mathbf{x}_1, k) g_{\nu_0}(\mathbf{x}, \mathbf{x}_1, k) d\Upsilon(\mathbf{x}_1) \tag{4.28}$$

where $g_{\nu_0}(\mathbf{x}, \mathbf{x}_1, k)$ is the Green's function when the refractive index throughout Υ_- is $\nu_0(\mathbf{x})$. The distorted wave Rayleigh-Gans (Born) approximation is predicated on approximating the total field, within the integral describing the scattered field, by the total field which exists within Υ_- when its refractive index is $\nu_0(\mathbf{x})$. The expression for the distorted wave field $\psi_{dw}(\mathbf{x}, k)$ is then

$$\psi_{dw}(\mathbf{x}, k) = \int_{\Upsilon_-} k^2(\nu_1^2(\mathbf{x}_1) - 1)\psi_{\nu_0}(\mathbf{x}_1, k) g_{\nu_0}(\mathbf{x}, \mathbf{x}_1, k) d\Upsilon(\mathbf{x}_1) \tag{4.29}$$

where $\psi_{\nu_0}(\mathbf{x}, k)$ is the scattered field due to $\nu_0(\mathbf{x})$. Since, in principle, all the quantities in the integrand are known, (4.29) is an explicit integral expression for the distorted wave field. The Green's function for $\nu_0(\mathbf{x})$ however, is not analytically available for an arbitrary function $\nu_0(\mathbf{x})$. In the generalised back-projection method, the Green's function for $\nu_0(\mathbf{x})$ is estimated by invoking the ray-tracing high-frequency approximation described in §3.1. The Green's function can then expressed as

$$g_{\nu_0}(\mathbf{x}, \mathbf{x}_1, k) \approx A_{\nu_0}(\mathbf{x}, \mathbf{x}_1) e^{jkS_{\nu_0}(\mathbf{x}, \mathbf{x}_1)} \tag{4.30}$$

where $A_{\nu_0}(\mathbf{x}, \mathbf{x}_1)$ and $S_{\nu_0}(\mathbf{x}, \mathbf{x}_1)$ satisfy the first transport equation (3.7) and eikonal equation (3.6), respectively, for $\nu_0(\mathbf{x})$. These quantities can be evaluated numerically by invoking any appropriate numerical method, such as ray tracing.

By invoking (4.30), the volume source formulation (3.28), and an incident field due to a point source, the distorted wave Rayleigh-Gans (Born) field $\psi_{dw}(\mathbf{x}, k)$ can be written as

$$\psi_{dw}(\mathbf{x}, k) = \int_{\Upsilon_-} \nu_1^2(\mathbf{x}_1) A_{\nu_0}(\mathbf{x}_s, \mathbf{x}_1) e^{jkS_{\nu_0}(\mathbf{x}_s, \mathbf{x}_1)} A_{\nu_0}(\mathbf{x}_1, \mathbf{x}) e^{jkS_{\nu_0}(\mathbf{x}_1, \mathbf{x})} d\Upsilon(\mathbf{x}_1) \quad (4.31)$$

where $\mathbf{x}_s$ denotes the position of the point source. (4.31) can be rewritten as

$$\psi_{dw}(\mathbf{x}, k) = \int_{\Upsilon_-} \nu_1(\mathbf{x}_1) a(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) e^{jkS(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x})} d\Upsilon(\mathbf{x}_1) \quad (4.32)$$

where

$$\begin{aligned} S(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) &= S_{\nu_0}(\mathbf{x}_s, \mathbf{x}_1) + S_{\nu_0}(\mathbf{x}_1, \mathbf{x}) \\ a(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) &= A_{\nu_0}(\mathbf{x}_s, \mathbf{x}_1) A_{\nu_0}(\mathbf{x}_1, \mathbf{x}) \end{aligned} \quad (4.33)$$

The temporal Fourier transform of (4.32) with respect to k is

$$\Psi_{dw}(\mathbf{x}, t) = \int_{\Upsilon_-} \nu_1^2(\mathbf{x}) a(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) \delta(t - S(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x})) d\Upsilon(\mathbf{x}_1) \quad (4.34)$$

which relates $\nu_1^2(\mathbf{x})$ to its integral along the rays implicit in $S(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) = \textit{constant}$. $\Psi_{dw}(\mathbf{x}, t)$ can be likened to the projection of $\nu_1(\mathbf{x})$ along each ray multiplied by the factor $a(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x})$. Since (4.34) is undefined for $t < 0$, Beylkin (1985a) calls it the causal generalised Radon transform. Inversion of (4.34) to reconstruct the discontinuities in the refractive index can be performed by invoking (Beylkin, 1985a)

$$R[\nu_1^2(\mathbf{x})] = \int_{\Upsilon_{obs}} \Psi_{dw}(\mathbf{x}_1, t)|_{t=S(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_1)} b(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) d\Upsilon(\mathbf{x}_1) \quad (4.35)$$

where R is here called the reconstruction operator and

$$b(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) = J(\mathbf{x}_1, \mathbf{x}) / a(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}) \quad (4.36)$$

where $J(\mathbf{x}_1, \mathbf{x})$ is the Jacobian of $S(\mathbf{x}_s, \mathbf{x}_1, \mathbf{x})$ with respect to $\mathbf{x}_1$ and $\mathbf{x}$. (4.35) is related to back-projection since the projections are projected back along lines of constant $S(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_1)$ (Beylkin, 1985a). The nature of the reconstruction operator R is discussed in the next paragraph.

Invoking pseudodifferential operator theory (Treves, 1980) it can be shown that $R = I + T$ where I is an operator which reconstructs $\nu_1(\mathbf{x})$ perfectly and T is a smoothing operator which operates on $\nu_1^2(\mathbf{x})$ such that it has at least one more continuous derivative than $\nu_1^2(\mathbf{x})$ (Beylkin, 1985a). Consequently, $R[\nu_1^2(\mathbf{x})]$ contains all the information about the discontinuities of $\nu_1^2(\mathbf{x})$ since adding a smooth function $T[\nu_1^2(\mathbf{x}]$ neither changes the location nor the size of a discontinuity. Thus R operates on the data to reconstruct or locate the discontinuities in a medium.

Beylkin (1985a) explicitly evaluates R for various $\nu_0(\mathbf{x})$ and particular source and receiver geometries. Schultz and Jaggard (1987) present numerical results derived from experimental data. They compare Beylkin's algorithm to classical back-projection and show that it represents a significant improvement. Levy and Esmersoy (1988) have extended the method in such a way that ray tracing is avoided.

4.2.2.5 Extended Rytov Approximations

The methods discussed in §4.2.2.1 through §4.2.2.4 are based upon either the Rayleigh-Gans (Born) or Rytov approximations. Research into devising more accurate methods has been active (Bates *et al.*, 1976; Soumekh, 1986; Devaney, 1986) but without notable success. This section reviews an approach, known as the the extended Rytov approximation, put forward by Bates *et al.* (1976) as a possible improvement to the Rytov approximation.

The Rytov approximation is presented in §3.2.2, where it is shown that the equivalent sources are approximated by $k^2(\nu^2(\mathbf{x})-1)\psi_i(\mathbf{x},k)$ on the assumption that $\nabla S_s(\mathbf{x})\cdot\nabla S_s(\mathbf{x})$ is negligible. The extended Rytov approximation attempts to obtain improved accuracy by partially compensating for the term $\nabla S_s(\mathbf{x})\cdot\nabla S_s(\mathbf{x})$. The ray-tracing high frequency approximation (see §3.1.1) to $\psi(\mathbf{x},k)$ is (refer to §3.1.1)

$$\psi(\mathbf{x},k)=\nu^{-\frac{1}{2}}(\mathbf{x})e^{-jk\int_l \nu ds} \tag{4.37}$$

where l denotes the ray from the source which passes through $\mathbf{x}$. When $\psi_i(\mathbf{x},k)=e^{j\mathbf{k}\cdot\mathbf{x}}$ the term $\nabla S_s(\mathbf{x})\cdot\nabla S_s(\mathbf{x})$ can be expressed as

$$\begin{aligned}\nabla S_s(\mathbf{x})\cdot\nabla S_s(\mathbf{x}) &= \nu^2(\mathbf{x})+1-2\nu(\mathbf{x})\hat{\mathbf{s}}\cdot\hat{\mathbf{k}}-\frac{j}{2k}\left(\hat{\mathbf{s}}\nu(\mathbf{x})-\hat{\mathbf{k}}\right)\cdot\nabla\ln\nu(\mathbf{x})\\ &\quad -\frac{1}{4k^2}(\nabla\ln\nu(\mathbf{x}))\cdot(\nabla\ln\nu(\mathbf{x}))\end{aligned} \tag{4.38}$$

where $\hat{\mathbf{s}}$ is the unit vector in the direction of the ray at $\mathbf{x}$. The only unknown in RHS (4.38) is $\hat{\mathbf{s}}\cdot\hat{\mathbf{k}}$. This can be approximated by $\hat{\mathbf{s}}\cdot\hat{\mathbf{k}}\approx 1$, which is only in error by 6 % when $\hat{\mathbf{s}}$ and $\hat{\mathbf{k}}$ differ by 20° (Bates *et al.*, 1976). The approximation is predicated on the assumption that the directions of the rays representing the total field do not deviate significantly from those representing the incident field. (4.38) can then be written as

$$\begin{aligned}\nabla S_s(\mathbf{x})\cdot\nabla S_s(\mathbf{x}) &= \nu^2(\mathbf{x})+1-2\nu(\mathbf{x})-\frac{j}{2k}(\nu(\mathbf{x})-1)\frac{\partial\ln\nu(\mathbf{x})}{\partial\hat{\mathbf{k}}}\\ &\quad -\frac{1}{4k^2}(\nabla\ln\nu(\mathbf{x}))\cdot(\nabla\ln\nu(\mathbf{x}))\end{aligned} \tag{4.39}$$

(3.42) can thus be approximately expressed as

$$\begin{aligned}(\nabla^2+k^2)\tilde{\psi}(\mathbf{x},k) &= -\left[2k^2(\nu(\mathbf{x})-1)+\frac{jk}{2}(\nu(\mathbf{x})-1)\frac{\partial\ln\nu(\mathbf{x})}{\partial\hat{\mathbf{k}}}\right.\\ &\quad \left.+\frac{1}{4}(\nabla\ln\nu(\mathbf{x}))\cdot(\nabla\ln\nu(\mathbf{x}))\right]\psi_i(\mathbf{x},k)\end{aligned} \tag{4.40}$$

If the RHS (4.40) is expressed as $\tilde{\sigma}(\mathbf{x})$ and the volume source formulation (3.28) is invoked, an integral expression relating $\nu(\mathbf{x})$ to $\tilde{\psi}(\mathbf{x},k)$ can be obtained (Bates *et al.*, 1976) as

$$\tilde{\psi}(\mathbf{x},k)=\int_{\Upsilon_-}\tilde{\sigma}(\mathbf{x}_1)\psi_i(\mathbf{x}_1,k)g(\mathbf{x},\mathbf{x}_1,k)d\Upsilon(\mathbf{x}_1) \tag{4.41}$$

If data are gathered throughout Υ_{obs} for a range of values $\mathbf{k}$ then (4.41) becomes a Fredholm integral equation of the first kind. Bates *et al.* (1976) suggest that (4.41) may then be inverted using the algebraic expansion technique presented in §2.2.8. The individual terms on the right hand side of (4.40) can be separated because of the known dependence on k expressed in (2.4).

Bates *et al.* (1976) also suggest that the most significant term on RHS (4.40) is $2k^2(\nu(\mathbf{x}) - 1)\psi_i(\mathbf{x}, k)$. It is therefore reasonable to approximate the density of the equivalent sources by $2k^2(\nu(\mathbf{x}) - 1)\psi_i(\mathbf{x}, k)$ than by either RHS (4.40) or by $k^2(\nu^2(\mathbf{x}) - 1)\psi_i(\mathbf{x}, k)$ as is assumed for the Rytov approximation. Intuitively it is expected that the incremental phase change is directly related to the product of conventional path length along a ray and $(\nu(\mathbf{x}) - 1)$. This also accords, of course, with (3.10). By invoking asymptotic techniques (Felson and Marcuvitz, 1973) for the extended Rytov approximation it can be shown that the incremental phase change per wavelength (see §2.2.3) along a ray is $k(\nu(\mathbf{x}) - 1)$. However, for the conventional Rytov approximation, the incremental phase change per wavelength is known to be $\frac{1}{2}k(\nu^2(\mathbf{x}) - 1)$, which does not agree with physical intuition. Consequently, the extended Rytov approximation appears to represent an improvement to the Rytov approximation in that it agrees more with physical intuition. However, the few numerical results which have so far been presented do not suggest that the improvements are likely to be spectacular (Dunlop *et al.*, 1976; Dunlop, 1978).

4.3 Implicit Solutions

When an explicit solution for the generalised constitutive parameter cannot be found, implicit solutions must be resorted to. A solution $\chi(\mathbf{x})$ is then sought subject to (4.1):

$$\psi(\mathbf{x}, k) = \Lambda(\chi(\mathbf{x})) \tag{4.42}$$

given $\psi(\mathbf{x}, k)$ throughout Υ_{obs}. The operator Λ is understood to provide either an exact or an approximate solution to the direct problem. It is also understood that the nature of the particular problem is such that Λ cannot be explicitly inverted. However, a number of difficulties tend to arise. The inverse of the operator Λ is usually ill-posed so that solutions are often either non-unique, non-existent, or depend discontinuously on the data. Also, because Λ is implicit and nonlinear, iterative methods of solution must be invoked.

4.3.1 Ill-posedness and Regularisation

Ill-posedness is a feature of many inverse problems (Bertero and Mol, 1981; Sleeman, 1982; Sabatier, 1983). An ill-posed operator can either make solutions of (4.1) unstable if noise is present or produce solutions which are incorrect. It is consequently important to be able to manipulate ill-posed operators into well-posed forms. Methods which permit this to be done are said to be regularisation techniques (Bertero and Mol, 1981).

A classic example of an ill-posed operator is that characterising the Fredholm integral equation of the first kind (see §2.2.8):

$$\int_a^b u(y)k(x, y)dy = h(x) \tag{4.43}$$

where $k(x, y)$ is a continuous kernel. Assuming that there exists a unique solution $u(x)$ corresponding to a given $h(x)$, it might seem to be impossible to alter $u(x)$ such that it is still a solution to (4.43). However, the Riemann-Lebesgue theorem confirms that

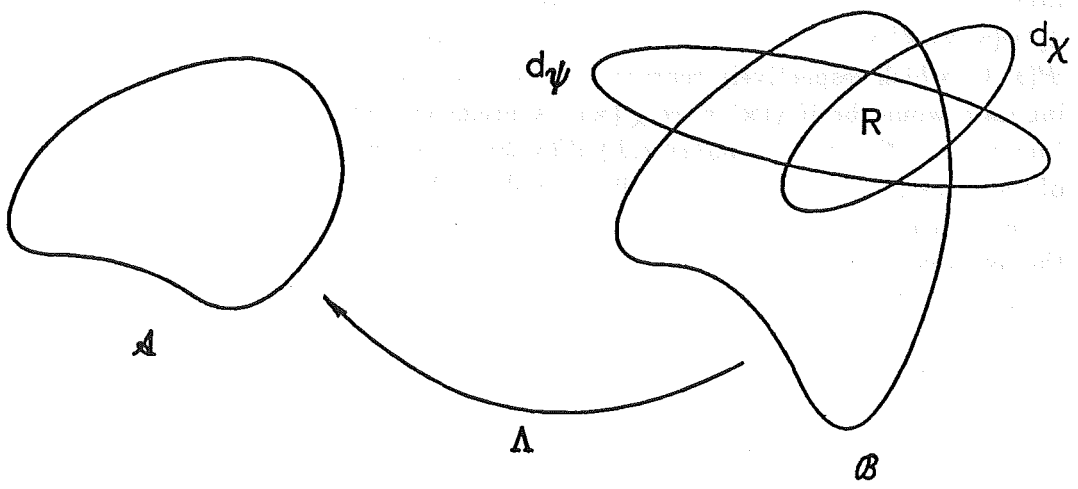


Figure 4.6: Graphical illustration of the mapping Λ. The operator Λ maps the function space $\mathcal{B}$ of all possible constitutive parameters to the space $\mathcal{A}$ of the resulting scattering data. The subsets $\mathbf{d}_\chi$ and $\mathbf{d}_\psi$ are generated after regularisation.

(Baltes, 1980)

$$\lim_{n\to\infty} C \int_a^b k(x,y)\sin(ny)dy = 0 \tag{4.44}$$

where C is an arbitrary constant and n is an integer. By taking C and n sufficiently large, and adding $C\sin(ny)$ to $u(x)$ in (4.43), widely different integrands are obtained without significantly altering $h(x)$. Consequently, the operator equation (4.43) is ill-posed in the sense that very small changes to $h(\mathbf{x})$ can change $u(\mathbf{x})$ greatly. There may also be several $u(\mathbf{x})$ corresponding to a given $h(\mathbf{x})$, which represents true non-uniqueness.

Methods of restoring an ill-posed operator can easily be understood graphically (Baltes, 1980; Bertero *et al.*, 1988), as is illustrated in fig. 4.6. The operator Λ maps the set of all constitutive parameters $\mathcal{B}$ to the set of fields $\mathcal{A}$ which result from the interaction of $\psi_i(\mathbf{x},k)$ with each constitutive parameter. Various regularisation methods are outlined in the paragraphs below.

Operator equations which do not provide unique solutions can generate many constitutive parameters $\chi(\mathbf{x})$ for scattering data $\psi(\mathbf{x},k)$ observed in Υ_{obs}. To contrive uniqueness, there must exist a criterion for choosing one of the solutions to be the most desirable (Nashed, 1981). This choice can be based on *a priori* information about the physics of the situation. Examples of such information are that the solution must exist within a specified location or that the solution must be real. One approach to choosing between solutions is then to minimize the distance $d_\chi(\chi_0(\mathbf{x}),\chi(\mathbf{x}))$ between a reference $\chi_0(\mathbf{x})$, which incorporates the *a priori* information, and $\chi(\mathbf{x})$. Graphically, incorporation of *a priori* information restricts the set of solutions $\mathcal{B}$ to a subset $\mathrm{d}\chi$, as illustrated in fig. 4.6.

Some operator equations cannot provide solutions for certain types of data. Solutions are then said to be non-existent. An operator exhibits non-existence of its

solutions when the scattering data lie outside the set $\mathcal{A}$ in which the operator is defined (Sabatier, 1983). This can be caused by the corrupting effects of noise. Regularisation can be invoked to force the operator to be well-posed in the following sense. Let $d_\psi(\psi(\mathbf{x},k),\psi'(\mathbf{x},k))$ be a measure of the distance between two functions $\psi(\mathbf{x},k)$ and $\psi'(\mathbf{x},k)$, which respectively represent the actual scattering data and what the scattering data would be if $\chi(\mathbf{x})$ were $\chi'(\mathbf{x})$. A preferred solution to the inverse problem is thus $\chi'(\mathbf{x}) \in \mathcal{B}$ such that $d_\psi(\psi(\mathbf{x},k),\psi'(\mathbf{x},k))$ is minimised, thereby restricting the set of solutions to a subset $\mathbf{d}_\psi$, as is illustrated graphically in fig. 4.6.

Regularisation can also restore continuity between the scattering data $\psi(\mathbf{x},k)$ and the solution $\chi(\mathbf{x})$ (Baltes, 1980). Both of the regularisation procedures described above ensure this, because they restrict the solution space to the set $\mathbf{R}$, which is the intersection of the spaces $\mathbf{d}_\chi$ and $\mathbf{d}_\psi$, as illustrated in fig. 4.6. Because the set $\mathbf{R}$ is a small subset of the set of all the possible solutions, it does not include solutions which are wildly different from the "exact" solution. This then avoids discontinuities between the scattering data and the solution (Baltes, 1980)

If the three regularistion techniques discussed above are combined then a general regularisation scheme is obtained for restoring ill-posed operator equations (Sabatier, 1983). This regularisation scheme can be expressed in general as

$$\min_{\chi}(1-\alpha)d_\chi(\chi_0,\chi) + \alpha d_\psi(\psi,\psi') \tag{4.45}$$

where $\alpha \in [0,1]$. Variation of α allows a certain trade off between the likelihood of $\psi(\mathbf{x},k)$ representing a useful solution and the quality of its data fit. While (4.45) characterises many regularisation schemes, d_ψ and $d\chi$ however are different in each scheme. The well known Tikhonov-Miller regularisation procedure can also be expressed in the above form (Sabatier, 1983).

4.3.2 Techniques for the Solution of Nonlinear Equations

There are two quite distinct approaches to obtaining a solution $\chi(\mathbf{x})$ to the implicit equation (4.42). In one approach (4.42) is treated as a nonlinear operator equation and in the other as a nonlinear system of algebraic equations.

4.3.2.1 Nonlinear Operator Method

The nonlinear operator method involves defining Λ as a nonlinear operator to act on the functions, $\chi(\mathbf{x})$ and $\psi(\mathbf{x},k)$ in a normed linear space. Existing mathematical techniques can then be invoked to derive iterative procedures for solving (4.42) (Rall, 1969).

The fixed-point algorithm is a general iterative procedure for solving nonlinear operator equations of the type (4.42) (Rall, 1969). One starts out by finding an appropriate rationale for rewriting (4.42) in the form

$$\chi(\mathbf{x}) = \Lambda^*(\chi(\mathbf{x})) \tag{4.46}$$

where Λ^* is a new operator which is derived from Λ and which depends on the scattering data $\psi_s(\mathbf{x},k)$. Given an initial estimate $\chi^{(0)}$ of the exact solution, an improved solution $\chi^{(1)}$ can then be generated iteratively as

$$\chi^{(l+1)}(\mathbf{x}) = \Lambda^*\left(\chi^{(l)}(\mathbf{x})\right) \tag{4.47}$$

where l is an integer and the $\chi^{(l)}$ are called iterates. The limit point of the iterates, if they in fact converge, is known as the fixed-point of Λ^* and it is a solution of (4.42) (Dennis and Schnabel, 1983). However, because the direct mapping is usually nonlinear with respect to $\chi(\mathbf{x})$, convergence of the fixed point algorithm is hard to establish.

The Newton-Kantorovich algorithm is based on the theory of the fixed-point algorithm for solving nonlinear operator equations (Rall, 1969). Once an initial estimate of $\chi^{(l)}$ is given, an improved estimate of the solution is obtained by setting

$$\chi^{(l+1)} = \chi^{(l)} + \delta\chi^{(l)} \tag{4.48}$$

where the perturbation $\delta\chi^{(l)}$ is found by linearising Λ about the current estimate of $\chi^{(l)}$ as

$$\Lambda\left(\chi^{(l)} + \delta\chi^{(l)}\right) \approx \Lambda\chi^{(l)} + \Lambda'(\chi^{(l)})\delta\chi^{(l)} \tag{4.49}$$

where Λ' denotes the Fréchet derivative of Λ (Rall, 1969). To compute an improved estimate $\chi^{(l+1)}$, the LHS (4.42) is equated to the RHS of (4.49) and the resulting equation

$$\Lambda'(\chi^{(l)})\delta\chi^{(l)} = \psi(\mathbf{x}, k) - \Lambda(\chi^{(l)}) \tag{4.50}$$

is solved for the update $\delta\chi^{(l)}$. The Newton-Kantorovich algorithm can be shown to converge quadraticly (Rall, 1969). Comparing (4.50) to (4.47) it is observed that the iterative scheme is in the form of the fixed-point algorithm. Whether or not the algorithm converges depends upon the faithfulness of the initial approximation $\chi^{(0)}$.

Many variants (such as the Levenberg-Marquadt algorithm) of the Newton-Kantorovich method have been developed (Dennis and Schnabel, 1983).

4.3.2.2 Nonlinear Algebraic Method

The nonlinear algebraic method defines $\chi(\mathbf{x})$ by a finite number of parameters rather than by a function. The inverse problem then reduces to determining these parameters from the scattering data. The nonlinear algebraic method has consequently become known as parameter identification (Parker, 1977).

The parameters are identified from the solutions to a nonlinear system of algebraic equations. Let χ_p be a vector representing the parameters and $\Lambda_p(\chi_p)$ be the solution of the direct problem with the given parameters χ_p. Then, if there are M parameters and N observations of $\psi(\mathbf{x}_n, k)$ for $n \in \{1, 2, \cdots N\}$ for $\mathbf{x}_n \in \Upsilon_{obs}$, a set of nonlinear algebraic equations is obtained as

$$0 = \min_{\chi_p} ||\Lambda_p(\chi_p) - \psi(\mathbf{x}_n, k)||_2 \quad n = 1, 2, \cdots N \tag{4.51}$$

In order to yield a unique solution, the number of parameters M must be less than or equal to the number of observations N.

The system of nonlinear equations may be solved for χ_p by standard optimisation techniques such as the Gauss-Newton, steepest descent or conjugate gradient methods (Fletcher, 1980). The Gauss-Newton method for a system of nonlinear equations, such as (4.51), is characterised by the following iterative scheme:

$$\chi^{(l+1)} = \chi_p^{(l)} + \delta\chi_p^{(l)} \tag{4.52}$$

where the update $\delta\chi_p^{(l)}$ satisfies

$$J^T(\chi_p^{(l)})J(\chi_p^{(l)})\delta\chi_p^{(l)} = J^T(\chi_p^{(l)})\left(\psi(\mathbf{x}_n,k) - \Lambda_p(\chi_p^{(l)})\right) \tag{4.53}$$

where $J(\chi_p^{(l)})$ is the Jacobian matrix of Λ_p (Fletcher, 1980).

In practice, the parameters are often determined by dividing the region Υ_- into a number of subregions, with the solution in each subregion being represented by some constant value, which has to be determined. The parameterisation can this be likened to discretisation.

The Newton-Kantorovich and Gauss-Newton algorithms are similar. Assuming that the same discretisation is invoked to solve the operator equations and algebraic equations, the resulting set of linear algebraic equations is the same for both methods. Discretisation of the problem followed by linearisation (algebraic method) gives the same result as linearisation and then discretisation (operator method) (Wouk, 1979; Connolly, 1989).

4.3.3 Review of Implicit Exact Solutions

This section reviews solutions to the inverse problem based on exact implicit formulas. The solutions discussed here are those presented by Roger (1981), Johnson and Tracy (1983), Colton (1984), Lesselier *et al.* (1985), Kristensson and Vogel (1986), and Colton and Monk (1988), all of whom support their conclusions with results of numerical simulations. These solutions are categorised according to the method by which the direct problem is solved.

4.3.3.1 Volume Source Formulation

The volume source formulation (see §3.2) can be incorporated into inverse methods for reconstructing the constitutive parameter $\chi(\mathbf{x})$ within Υ_- (Johnson and Tracy, 1983; Johnson *et al.*, 1984; Lesselier *et al.*, 1985). The approach taken here is that of parameter identification §4.3.2.2. For this purpose the volume source formulation is expressed as in (3.28):

$$\begin{aligned}\psi_s^{[m]}(\mathbf{x},k) &= \psi^{[m]}(\mathbf{x},k) - \psi_i^{[m]}(\mathbf{x},k) \\ &= \int k^2(\chi^2(\mathbf{x}_1,k)-1)\psi^{[m]}(\mathbf{x}_1,k)g(\mathbf{x},\mathbf{x}_1,k)d\Upsilon(\mathbf{x}_1)\end{aligned} \tag{4.54}$$

where the superscript $[m]$ relates to an incident field which is a plane wave whose direction is the m^{th} out of a total of M different specified directions.

Numerical inversion of (4.54) is eased by first expanding $(\chi^2(\mathbf{x})-1)\psi^{[m]}(\mathbf{x},k)$ as the series

$$(\chi^2(\mathbf{x})-1)\psi^{[m]}(\mathbf{x},k) = \sum_{n=1}^{N}(\chi^2(\mathbf{x}_n)-1)\psi^{[m]}(\mathbf{x}_n,k)b(\mathbf{x}-\mathbf{x}_n) \tag{4.55}$$

where $\mathbf{x}_n$ is the n^{th} point on a rectangular grid having a total of N points and $b(\cdot)$ is a suitable basis function, for example a rectangular or sinc function. Upon substituting the series expansion (4.55) into the volume source formulation (4.54), the scattered

field is expressed as

$$\psi_s^{[m]}(\mathbf{x}_p,k)=\psi^{[m]}(\mathbf{x}_p,k)-\psi_i^{[m]}(\mathbf{x}_p,k)=\sum_{n=1}^{N}(\chi(\mathbf{x}_n)^2-1)\psi^{[m]}(\mathbf{x}_n,k)C_{n,p} \tag{4.56}$$

where

$$C_{n,p}=\int_{\Upsilon_-} b(\mathbf{x}_1-\mathbf{x}_n)g(\mathbf{x}_p,\mathbf{x}_1,k)d\sigma(\mathbf{x}_1) \tag{4.57}$$

where $\mathbf{x}_p$ is the p^{th} point in a set of P observation points. (4.56) is an implicit expression for $\chi(\mathbf{x})$, which can be solved using optimisation techniques (see §4.3). Because the constitutive parameter is discretised before optimisation, nonlinear algebraic equations have to be solved. Methods for solving this type of equation are presented in §4.3.2.2.

Johnson and Tracy (1983) present the results of a numerical solution to (4.56) using a sinc function as the basis function $b(\cdot)$. This implies that the equivalent source density must be bandlimited to yield an accurate series representation (4.55). They use an optimisation technique known as the "alternating variable linear Kacmar method of optimisation", which can be summarised as (Johnson and Tracy, 1983)

1) Guess $(\chi^2(\mathbf{x})-1)$ and $\psi^{[m]}(\mathbf{x},k)$ in Υ_-

2) Ensure that the equivalent sources $(\chi^2(\mathbf{x})-1)\psi^{[m]}(\mathbf{x},k)$ generate the guessed $\psi^{[m]}(\mathbf{x},k)$. This is effected by holding $(\chi^2(\mathbf{x})-1)$ fixed and solving for $\psi^{[m]}(\mathbf{x},k)$ inside Υ_-

3) Using $\psi^{[m]}(\mathbf{x},k)$ from step 2, calculate $\chi(\mathbf{x})$ using the scattering data $\psi_s^{[m]}(\mathbf{x},k)$ and (4.56) in Υ_{obs}.

4) If $|\psi^{[m]}(\mathbf{x},k)-\psi^{(l)}(\mathbf{x},k)|_1<\epsilon$, where (l) refers to the l^{th} iteration and ϵ characterises the desired accuracy of the results, then stop. Otherwise go back to step 2.

It should be noted that, for each iteration, $\psi^{[m]}(\mathbf{x},k)$ must be calculated throughout both Υ_- and Υ_{obs}.

Johnson and Tracy (1983) present numerical simulations of reconstructions of the two-dimensional complex refractive index distribution

$$\chi(\mathbf{x})=1.0+Ae^{-\frac{2\rho^2}{\lambda^2}} \tag{4.58}$$

where A is a complex constant and ρ is the radial distance form the origin of the distribution. The reconstructions are computed on a 11×11 rectangular grid having a spacing of $\frac{1}{4}\lambda$. The maximum values of the parameters m and p in (4.56) are 17. This corresponds to overdetermining the system of equations in (4.56) by 140 %. When $A=(0.2+j0.02)$, Johnson and Tracey find that, after 20 iterations, the relative one norm error (see §2.2.9) between the original and reconstructed distributions is 10^{-3}. With 2% Gaussianly distributed noise (see §2.2.11) added to the scattered field, the relative one norm error is 0.035 after 20 iterations. Johnson and Tracy (1983) also investigate the conditioning and accuracy of the method under different conditions. They indicate that for a given grid-spacing the method becomes better conditioned as the value of A is reduced. They also show that the location of Υ_{obs} has little effect on the conditioning when the system is overdetermined by more than 300 %.

Lesselier *et al.* (1985) also employ the volume source formulation to reconstruct a complex refractive index distribution. They adopt a rectangular function for the basis function $b(\cdot)$. Their optimisation algorithm is based on the conjugate-gradient method, which is described in many texts (cf. Fletcher, 1980). They present simulated reconstructions of two objects on 30×30 rectangular grids. The first reconstruction is of a circular cylinder consisting of two concentric shells having different refractive indexes:

$$\chi(\mathbf{x}) = \begin{cases} 0.82 & \text{for} \quad 0 < \rho \leq 0.9\lambda \\ 0.80 & \text{for} \quad 0.9\lambda < \rho \leq 3\lambda \\ 1.0 & \text{otherwise} \end{cases} \tag{4.59}$$

After 20 iterations the RMS error between the original object and the reconstruction is 10^{-3}. The other reconstruction is of a simulated cross-section of a human neck the refractive index of which is written as

$$\chi(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{for} \quad 0 < \rho \leq 0.1\lambda \\ 1.0 & \text{otherwise} \end{cases} \tag{4.60}$$

where $f(\mathbf{x})$ simulates the distribution of refractive index over a typical human neck cross-section, such that $0.3 < f(\mathbf{x}) < 1.0$. After 20 iterations the RMS error is 10^{-2}. Lesselier *et al.* (1985) note that the stability of their method improves as the range, over which $f(\mathbf{x})$ varies decreases.

4.3.3.2 Null-field Method

The null-field method (see §3.3) can be implemented as an inverse technique for determining the shape of an object on the surface of which either Dirichlet or Neumann boundary conditions are satisfied. The technique is here formulated in two dimensions, for expositional convenience. An implicit relation between the shape $\sigma(\mathbf{x})$ of a cylinder of arbitrary cross-section and the scattering data is obtained. This relation can be optimised to yield a reconstruction of $\sigma(\mathbf{x})$ by any of the nonlinear operator methods described in §4.3.2.1 (Wall *et al.*, 1985; Kristensson and Vogel, 1986).

Inversion of the null-field method is an ill-posed procedure (Colton, 1984). Consequently, when the null-field method is invoked for determining the shape of a cylinder of arbitrary cross-section $\sigma(\mathbf{x})$, regularisation as described in §4.3.1 is required. Thus, nonlinear operator optimisation is invoked to determine $\sigma(\mathbf{x})$ by solving

$$L(\sigma) = \min_{\sigma} [d_{\psi}(\Lambda(\sigma) - \psi) + \alpha d_{\chi}(\sigma)] \tag{4.61}$$

where Λ represents the null-field solution of the direct problem α is a fixed parameter and d_{χ} and d_{ψ} are the regularising functions. The problem is then discretised. If $\sigma(\mathbf{x})$ is assumed to be star-like (i.e. single valued with respect to the angle θ) it can be expressed in a Fourier series as

$$\sigma(\theta) = a_0 + \sum_{m=1}^{\infty} a_m \cos(m\theta) + b_m \sin(m\theta) \tag{4.62}$$

where the a_m and b_m are constant expansion coefficients.

Kristensson and Vogel (1986) report simulated results of reconstructing the shape of the cross-section of a cylinder on the surface of which Dirichlet boundary conditions

are satisfied. They invoke the null-field method and the penalised likelihood method for optimisation and regularisation. This implies that the regularising functions are expressed as

$$\begin{aligned} d_\psi(\Lambda(\sigma) - \psi(\mathbf{x}, k)) &= ||\Lambda(\sigma) - \psi(\mathbf{x}, k)||^2 \\ d_\chi(\sigma) &= \int_0^{2\pi} \sigma(\mathbf{x})^2 d\sigma(\mathbf{x}) \end{aligned} \tag{4.63}$$

and the quasi-Newton method (Fletcher, 1980) is invoked to solve (4.61). Kristensson and Vogel (1986) use finite differences to approximate the Fréchet derivatives of $L(\sigma)$ needed in the quasi-Newton method. This method can be numerically expensive because the direct problem has to be solved twice to calculate each Fréchet derivative.

The overall implementation of the algorithm comprises four steps:

1) α is initially set to 10^{-1} and $\sigma(\mathbf{x})$ is initially approximated as a circle of diameter R.

2) An improved estimate of $\sigma(\mathbf{x})$ is obtained using the quasi-Newton method (see §4.3.2.1) to minimise (4.61)

3) The iterations are stopped if $||\Lambda(\sigma) - \psi(\mathbf{x}, k)||^2 \approx s^2$, where s^2 is the variance of the noise. Otherwise, one goes to step 4.

4) α is decremented by multiplying it by 10^{-1} and one returns to step 2.

Kristensson and Vogel (1986) report numerical examples for 7 objects whose shapes are defined by the canonical equation

$$\sigma(\theta) = \left[\left(\frac{2\pi \cos\theta}{a\lambda} \right)^{|\gamma|} + \left(\frac{2\pi \sin\theta}{b\lambda} \right)^{|\gamma|} \right]^{-\frac{1}{\gamma}} \tag{4.64}$$

where γ assumes a different value for each object. The values of a and b range between 1 and 5. The incident field is planar and the scattering data are specified in the far-field at nine equally spaced angles in the range 0° to 320°. The scattering data are in the form of either complex amplitude, magnitude only or total cross-section, which quantities are defined in §2.2.10. Numerical results for some of the objects are summarised in table 4.1.

4.3.3.3 Surface Integral Equation Method

The surface integral equation method (see §3.4) can also be implemented as an inverse technique for determining the shape of an object on the surface of which either Dirichlet or Neumann boundary conditions are satisfied (Roger, 1981; Colton, 1984). Because the surface integral and null-field methods are closely related, (4.61) and (4.62) can also be invoked to express the inverse problem as an implicit equation (Colton, 1984). An advantage of the surface integral technique is that the Fréchet derivatives can be represented as Fredholm integral equations (Roger, 1981). This overcomes difficulties

Object			Data type	Incident fields		Error		
γ	a	b	(see §2.2.10)	Range of directions	Number	ϵ_1	ϵ_2	ϵ_3
6	4	2	complex	0° – 270°	4	0.01	0.03	0.05
-2	2	1	complex	0° – 270°	4	0.01	0.07	0.09
2	5	2	complex	0° – 0°	1	0.02	-	-
2	5	2	complex	90° – 90°	1	0.15	-	-
2	5	2	complex	0° – 90°	2	0.015	-	-
2	5	2	complex	0° – 180°	2	0.015	-	-
2	5	2	complex	0° – 270°	4	0.01	-	-
2	5	2	amplitude	0° – 270°	4	0.01	-	-
2	4	2	cross	0° – 355.2°	75	0.01	0.05	0.12

Table 4.1: Summary of the performance of the Kristensson and Vogel (1986) algorithm. The object is specified according to (4.62). The errors ϵ_1, ϵ_2 and ϵ_3, which are specified after 5 iterations as relative norm 2 errors (see §2.2.9), are respectively for uncontaminated data and data corrupted with 1% and 10% Gaussianly distributed noise (see §2.2.11).

associated with having to calculate such derivatives using finite differences, as is done by Kristensson and Vogel (1986).

Colton (1984) reports numerical results pertaining to reconstructing the cross-sectioned shapes of cylinders on the surfaces of which Dirichlet boundary conditions are satisfied. The regularisation function d_ψ is expressed as

$$d_\psi(\Lambda(\sigma) - \psi(\mathbf{x}, k)) = ||\Lambda(\sigma) - \psi(\mathbf{x}, k)|| \tag{4.65}$$

The regularistion function d_χ is implemented by restricting χ to belong to the set of σ defined by $a \leq \sigma(\theta) \leq b$ and $|\sigma''(\theta)| \leq C$, where a, b and C are arbitrary constants. Optimisation of (4.61) is effected with a Newton method due to Madsen and Schjaer-Jacobsen (1978). The Fréchet derivatives are found as solutions to Fredholm integral equations.

Colton (1984) presents numerical solutions for two objects. The scattering data are in the form of the total cross-section (see §2.2.10). Colton's results are summarised in table 4.2. Roger (1981) also examines the problem of determining the shape of a cylinder on which Dirichlet boundary conditions are satisfied. He invokes the Newton-Kantorovich algorithm and a method similar to Colton's (1984) for solving the direct problem. He incorporates Tikhonov-Miller regularisation and obtains the Fréchet derivatives as solutions to Fredholm integral equations.

Roger (1981) presents numerical results, summarised here in table 4.3, for 6 objects. The scattering data are specified as magnitude (see §2.2.10) and are contaminated with 0.01% Gaussian distributed noise (see §2.2.11). Table 4.3 summarises these results. Since Roger (1981) only compares his original objects and his reconstructions of them by displaying them pictorially, the quantitative accuracy of his reconstructions can only be estimated. Each estimate is calculated by approximately evaluating the difference

Object	Incident fields		Parameters			Error	
	Range of directions	Number	a	b	C	ϵ	Number of iterations
$(\frac{\pi x}{2\lambda})^2 + (\frac{3\pi y}{\lambda})^2 = 1$	$0° - 90°$	10	0.2	∞	10	0.0055	3
$(\frac{2\pi x}{\lambda})^{10} + (\frac{4\pi y}{3\lambda})^{10} = 1$	$0° - 180°$	19	0.8	∞	20	1.179	4
$(\frac{2\pi x}{\lambda})^{10} + (\frac{4\pi y}{3\lambda})^{10} = 1$	$0° - 180°$	19	1.0	∞	100	1.340	3

Table 4.2: Summary of the performance of the Colton (1984) algorithm. The scattering data are specified in the form of total cross-section (see §2.2.10). The error is specified as the value of $L(\sigma)$, as defined by (4.61), once the algorithm has terminated. All the examples are for noiseless scattering data.

Object	Incident fields		Error
	Range of directions	number	
$\sigma(\theta) = \frac{\lambda}{\pi} - \frac{\lambda}{2\pi}\cos 2\theta$	$0° - 0°$	1	0.0
$\sigma(\theta) = \frac{\lambda}{2\pi} - \frac{\lambda}{4\pi}\cos 2\theta$	$0° - 0°$	1	0.08
$\sigma(\theta) = \frac{2\lambda}{\pi} - \frac{\lambda}{\pi}\cos 2\theta$	$0° - 0°$	1	0.2
$\|\|\mathbf{x}\|\|_\infty = \frac{0.1\lambda}{\pi}$	$0° - 360°$	many	0.05
$\|\|\mathbf{x}\|\|_\infty = \frac{0.05\lambda}{\pi}$	$0° - 360°$	many	0.13
$\|\|\mathbf{x}\|\|_\infty = \frac{0.4\lambda}{\pi}$	$0° - 360°$	many	0.01

Table 4.3: Summary of the performance of the Roger (1981) algorithm. The scattering data are specified in the form of total cross-section (see §2.2.10). The results listed in this table are those obtained after 5 iterations and with 1% additive Gaussian noise (see §2.2.11) added to the scattering data. The error measure is explained in the text.

between the area of the reconstruction and the original object and dividing it by the area of the original object.

4.3.3.4 Herglotz Wave Function Method

Colton and Monk (1985, 1986, 1987, 1988) have developed a method for reconstructing spatially inhomogeneous distributions of a constitutive parameter from far-field scattering data over a range of frequencies. The method stems from a technique for determining the shape of the cross-section of a cylinder, on whose surface Dirichlet boundary conditions are satisfied, from far-field scattering data specified over ranges of frequencies and incident and scattered directions (Colton and Monk, 1985). Kirsch *et al.* (1988) compare this method with Kirsch and Kress's method (see §4.4.4). Ochs (1987) has applied Colton and Monk's method to situations in which the scattering data are limited to small angular ranges. Colton and Monk's (1985) method relies on the Herglotz wave function, which is a solution to the Helmholtz equation (2.5) of the form

$$f(\rho;\phi,k) = \int_0^{2\pi} h(\theta) e^{jk\rho\cos(\theta-\phi)} d\theta \tag{4.66}$$

where $h(\theta)$ is known as the Herglotz kernel.

The application of (4.66) to the inverse scattering problem is conveniently illustrated by considering scattering from a surface $\sigma(\mathbf{x})$ on which the Dirichlet boundary condition is satisfied. If k is then an eigenvalue of the Laplacian in Υ_-, Herglotz wave functions can be found such that (Colton and Monk, 1985)

$$f(\sigma(\theta),k) = 0 \tag{4.67}$$

where the surface $\sigma(\mathbf{x})$ is taken to be star-like (i.e. single valued with respect to θ) so that it can be parameterised as $\sigma(\theta)$. (4.67) is simply a statement of the need to impose the Dirichlet boundary condition on $\sigma(\mathbf{x})$.

Due to the condition (4.67), the Herglotz kernel is orthogonal to the far-field pattern (Colton and Monk, 1985), which can be demonstrated by expressing the far-field pattern as reradiation from sources on $\sigma(\mathbf{x})$. Invoking (3.45) and (2.19) leads to

$$\psi_{ff_s}(\theta,k) = \int_\sigma \frac{\partial\psi(\rho;\phi)}{\partial\mathbf{n}} e^{jk\rho\cos(\theta-\phi)} d\sigma(\rho;\phi) \tag{4.68}$$

If both sides of (4.68) are multiplied by the Herglotz kernel, which is introduced in (4.66), and integrated with respect to θ, then

$$\int_0^{2\pi} h(\theta)\psi_{ff_s}^{[m]}(\theta,k)d\theta = 0 \tag{4.69}$$

where $\psi_{ff_s}^{[m]}(\theta,k)$ is the far-field scattering pattern corresponding to an incident plane wave whose direction of propagation is the m^{th} belonging to a set of M different directions. (4.69) confirms that the far-field pattern is orthogonal to the Herglotz kernel in (4.67).

A solution to this inverse problem of determining the surface $\sigma = \sigma(\theta)$ can thus be expressed as

$$L(\sigma) = \min_{\sigma,h,k} \sum_{m=1}^{M} \left| \int_0^{2\pi} h(\theta)\psi_{ff_s}^{[m]}(\theta,k) \right|^2 + \int_0^{2\pi} |f(\theta)|^2 d\theta \tag{4.70}$$

Both terms on RHS (4.70) vanish when the correct form for $\sigma(\mathbf{x})$ is found because of (4.67) and (4.69). In practice, because the eigenvalue k is not known *a priori*, scattering data over a range of k are therefore required. However, it is known that, for a surface bounded by a and b such that

$$a \leq \sigma(\theta) \leq b \tag{4.71}$$

the first eigenvalue of the Laplacian is somewhere in the interval (Ochs, 1987)

$$\left[\frac{k_0^2}{b^2}, \frac{k_0^2}{a^2}\right] \tag{4.72}$$

where k_0 denotes the first zero of the Bessel function of the first kind of order zero. Consequently, the frequency range, over which scattering data must be specified, can be deduced.

Object	Number of incident fields	Number of data points	Error
$\left(\frac{2\pi x}{\lambda}\right)^2 + (\frac{4\pi y}{3\lambda})^2 = 1$	13	13	0.001
$\sigma(\theta) = \frac{\lambda}{2\pi} + \frac{\lambda}{4\pi}\cos\theta$	15	17	0.0148
$\sigma(\theta) = \frac{\lambda}{2\pi} - \frac{\lambda}{4\pi}\cos 2\theta$	17	21	0.0815
$(\frac{2\pi x}{\lambda})^{10} + (\frac{4\pi y}{3\lambda})^{10} = 1$	15	17	0.0441

Table 4.4: Summary of the performance of the Colton and Monk (1986) algorithm. The scattering data are specified in the form of complex amplitudes (see §2.2.10), the incident field illuminates the object from equally spaced angles in the range 0° to 360° and the error is specified as the relative norm 2 error.

Colton and Monk's approach is noteworthy because the interior eigenfunctions of the unknown surface $\sigma(\mathbf{x})$ are linked to the far-field pattern by the kernel $h(\theta)$. This means that it is unnecessary to solve an integral equation at each iteration, so that a computationally efficient algorithm can be devised. Uniqueness and continuity of $\psi(\mathbf{x}, k)$ on $\sigma(\mathbf{x})$ are also considered in detail by Colton and Monk (1985) as is regularisation. They invoke a Newton method for the minimisation of (4.70).

Colton and Monk (1986) present many numerical results. Those for reconstructing shapes $\sigma(\mathbf{x})$ on which Dirichlet boundary conditions are satisfied are listed in table 4.4. In later papers they present results for penetrable homogeneous and inhomogeneous objects (Colton and Monk; 1986; 1987; 1988).

4.3.4 Review of Implicit Approximate Solutions

In many situations, the linear dimensions of the objects to be reconstructed are much larger than the wavelength of the emanations. Exact implicit formulations then become computationally intractable. It is consequently necessary to use approximate formulations of scattering to reduce the computational load. These formulations do not, however, necessarily lead to explicit relationships between the constitutive parameter and the scattering data. Thus, implicit solution techniques are also needed. In the following sections approximate implicit solutions to the inverse problem are described. They are again classified according to the method invoked for solving the direct problem.

4.3.4.1 Ray Tracing

There are many implementations of methods based on ray tracing (Schomberg, 1978; Lytle and Dines, 1980; McKinnon and Bates, 1980; Cha and Vest, 1981; Peterson *et al.*, 1988; Lo *et al.*, 1988). Because ray-tracing is a high frequency approximation, it allows objects whose linear dimensions are large compared to the wavelength to be reconstructed reasonably accurately and computationally efficiently.

In ray-tracing, the relationship between the constitutive parameter and the scatter-

ing data is described through the eikonal equation, which is, with reference to (3.6)

$$\nabla S(\mathbf{x}) \cdot \nabla S(\mathbf{x}) = \nu^2(\mathbf{x}) \tag{4.73}$$

where $S(\mathbf{x})$ is the path length of a ray from the source to a point $\mathbf{x}$. Inverse methods based on ray-tracing involve inverting (4.73). All of the methods published to date implement this by expressing (4.73) equivalently as the minimised line integral (3.10), which is repeated here for convenience:

$$S(\mathbf{x}) = \min_{path} \int_{s_0}^{\mathbf{x}} \nu(\mathbf{x})\, ds \tag{4.74}$$

where $\mathbf{s}_0$ represents the source point. This equation is nonlinear because the ray paths over which $\nu(\mathbf{x})$ is integrated depend not only on the distance between $\mathbf{s}_0$ and $\mathbf{x}$ but also on $\nu(\mathbf{x})$.

The quantity $S(\mathbf{x})$ in (4.74) can be likened to a projection of $\nu(\mathbf{x})$ along the ray paths from $\mathbf{s}_0$ to $\mathbf{x}$. Therefore, the geometry and terminology used for CT (see §4.2.1.1) is also suitable here. Fig. 4.7 shows two systems (x, y) and (ξ, η) of Cartesian coordinates having the common origin O with (ξ, η) being rotated by ϕ with respect to (x, y). The dashed line joining the source and sink in fig. 4.7 is a ray which is released initially parallel to the η axis and terminates at the sink whose position is specified by ξ along the ξ-axis. The ray path which terminates at the particular sink, identified by the symbol (ξ, ϕ), is denoted by $R_{\xi,\phi}(\nu)$. The data relating to each sink represent the values of $S(\mathbf{x})$ along rays between the sink and the various sources. These data can be treated as projections $p(\xi; \phi)$ of $\nu(\mathbf{x})$ along the rays $R_{\xi,\phi}(\nu)$. For sufficiency $p(\xi; \phi)$ must be specified for a range of angles ϕ. In terms of this notation, (4.74) can be rewritten as

$$p(\xi; \phi) = \int_{R_{\xi,\phi}(\nu)} \nu(\mathbf{x})\, dl \tag{4.75}$$

The inverse problem can then be posed as: reconstruct $\nu(\mathbf{x})$ from all the given projections $p(\xi; \phi)$. The difficulty, however, is that the ray paths are initially unknown within Υ_- because the spatial variations of $\nu(\mathbf{x})$ are themselves initially unknown.

An approach which enables $\nu(\mathbf{x})$ to be reconstructed is based on initially approximating all the $R_{\xi,\phi}(\nu)$ by straight rays. The integral (4.75) then represents projections of the quantity $\nu(\mathbf{x})$, which can be reconstructed using the CT techniques described in §4.2.1.1, as indicated by Glover and Sharp (1977), Mueller *et al.* (1979) and Peterson *et al.* (1988). McKinnon and Bates (1980) and Cha and Vest (1981) present algorithms which iteratively improve this straight ray approximation, by invoking the following sequence of steps:

1) An estimate of $\nu(\mathbf{x})$ is reconstructed from the given projections $p(\xi; \phi)$ assuming straight rays.

2) The rays through the estimate of $\nu(\mathbf{x})$ are traced.

3) Projections $p_e(\xi, \phi)$, defined by (4.75), are computed for this estimate of $\nu(\mathbf{x})$.

4) Straight ray projections $p_s(\xi, \phi)$ are computed for this estimate of $\nu(\mathbf{x})$.

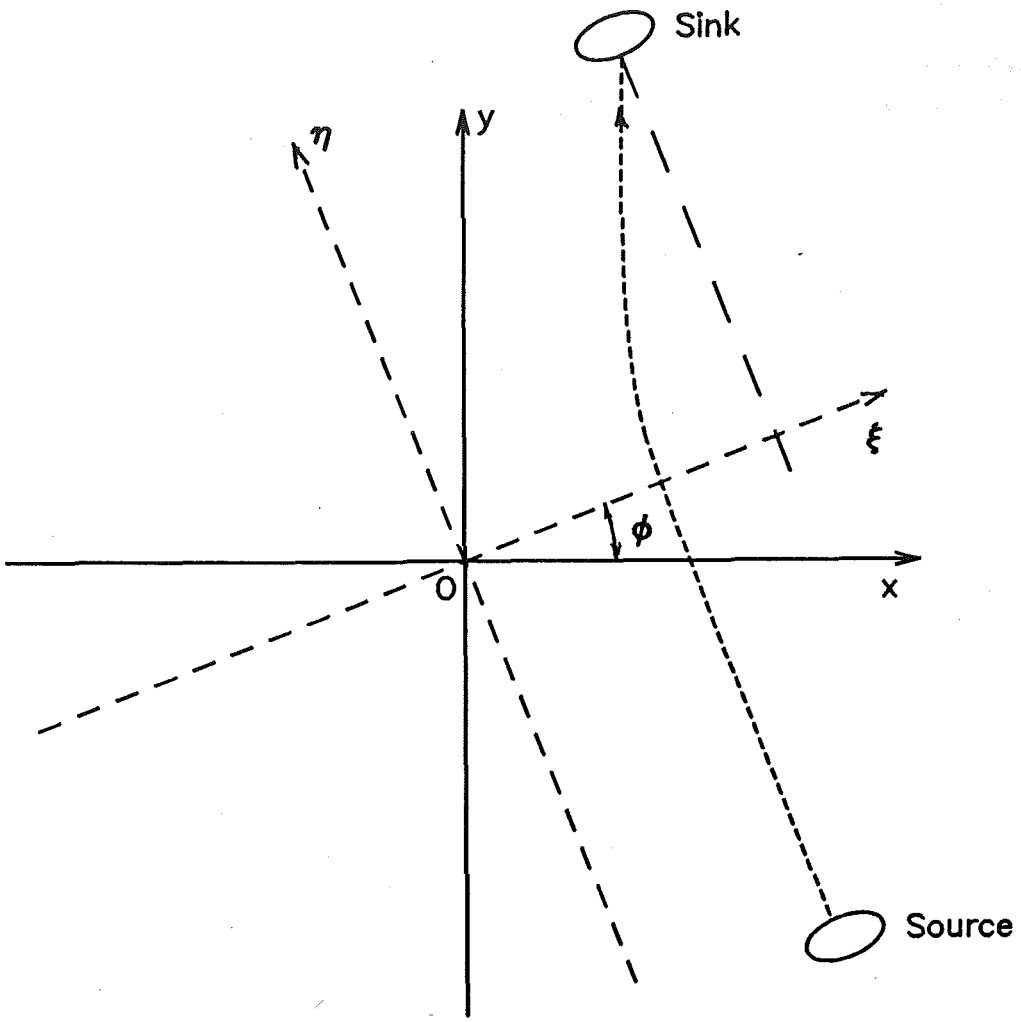


Figure 4.7: Coordinate systems and notational details required for solving the inverse problem using ray-tracing in two dimensions. The dashed line represents a ray which is released from the source, initially in a direction parallel to the η-axis, and which terminates at the sink.

5) Difference projections $D(\xi;\phi) = [p_s(\xi;\phi) - p_e(\xi;\phi)]$ are computed.

6) The given projections are replaced by the corrected projections $[p(\xi;\phi) + D(\xi;\phi)]$.

7) Step 1) is repeated. The iterations are stopped if, throughout Υ_-, the new $\nu(\mathbf{x})$ differs from the previous $\nu(\mathbf{x})$ by less than a threshold corresponding to the desired accuracy for the reconstruction of $\nu(\mathbf{x})$. Otherwise, one proceeds to step 2).

McKinnon and Bates (1980) present results of reconstructions of radially symmetric distributions. Improvements over straight ray reconstructions are obtained for refractive index distributions which deviate by up to 50% from their average values. Cha and Vest (1981) present reconstructions of both symmetric and asymmetric refractive index distributions. Improvements over straight ray reconstructions are demonstrated for deviations of up to 70% from their average values, with the infinity norm (see §2.2.9) for the difference between the original and reconstructed versions of each $\nu(\mathbf{x})$ never exceeding 1% of the average value of $\nu(\mathbf{x})$.

Another viable approach to reconstructing refractive index distributions is to account for the ray curvature directly, as demonstrated with the aid of iterative algebraic techniques by Schomberg (1978), Lytle and Dines (1980), Peterson *et al.* (1988) and Lo *et al.* (1988). The scattering region Υ_- is first divided into M rectangles. In the m^{th} rectangle the refractive index is denoted by ν_m. Projections, as defined by (4.75), are given for N rays spanning appropriate ranges of ξ and ϕ. The value of RHS (4.75) is denoted p_n for the n^{th} ray. The length of the n^{th} ray passing through the m^{th} rectangle is denoted by $l_{m,n}$. Then

$$p_n = \sum_{m=1}^{M} l_{m,n}\nu_m \quad n = 1 \cdots N \tag{4.76}$$

To invert (4.76), so as to estimate ν_m for each m, an iterative approach is adopted. The algorithm starts with an initial estimate of each ν_m denoted by $\nu_m^{(0)}$. Initial estimates of the p_n can then be calculated with the aid of the formula

$$p_n^{(0)} = \sum_{m=1}^{M} l_{m,n}^{(0)}\nu_m^{(0)} \quad n = 1 \cdots N \tag{4.77}$$

after the $l_{m,n}^{(0)}$ have been calculated by ray tracing through the estimated refractive index distribution characterised by the $\nu_m^{(0)}$.

The iterative algorithm repeatedly updates the initial estimates $\nu_m^{(0)}$ by invoking the recurrence formula known as the algebraic reconstruction technique (ART) (Herman, 1980, Chapter 11):

$$\begin{aligned} \nu_m^{(l+1)} &= \nu_m^{(l)} + \nu_{m,n}^{(l)} \\ &= \nu_m^{(l)} + l_{m,n}^{(l)} \frac{p_n - \sum_{m=1}^{M} l_{m,n}^{(l)}\nu_m^{(l)}}{\sum (l_{m,n}^{(l)})^2} \end{aligned} \tag{4.78}$$

where $\nu_{m,n}^{(l)}$ is the correction of $\nu_m^{(l)}$ generated, in the manner indicated above, by examining the n^{th} ray. (4.78) is a least squares solution of $p_n - p_n^{(l)} = 0$.

Object	Number of incident fields	Number of data points	Error
$x(t) = \frac{\lambda}{2\pi}\cos t + \frac{\lambda 0.65}{2\pi}(\cos 2t - 1)$ $y(t) = \frac{3\lambda}{4\pi}\sin t \quad t \in [0, 2\pi]$	12	32	0.187
$\left(\frac{2\pi x}{\lambda}\right)^{10} + \left(\frac{4\pi y}{3\lambda}\right)^{10} = 1$	12	32	0.036

Table 4.5: Summary of the performance of the Kirsch and Kress (1987) algorithm. The scattering data are specified in the form of complex amplitudes (see §2.2.10), the incident field illuminates the object from equally spaced directions in the angular range 0° to 360° and the error is specified as the relative norm 2 error.

A variation of this algorithm can be invoked. It is known as the simultaneous iterative reconstruction technique (SIRT) (Gilbert, 1972). It is similar to ART except that the correction to each ν_m is determined from all the p_n simultaneously. Lytle and Dines (1980) illustrate this method, after regularisating it, with the results of computer simulations. Visually, their method generates reconstructions which seem to be significant improvements over the straight ray reconstructions for refractive index distributions exhibiting deviations of up to 30% from their average values, when 1% Gaussian noise (see §2.2.11) is added to each p_n. Schomberg's (1978) results are not as visually impressive, mainly because the maximum deviation of the refractive index distribution from its average is only 5%, so that the straight ray approximation is reasonably accurate anyway. Lo *et al.* (1988) present results for this method which they also compare with results of Rayleigh-Gans (Born) inversion of data measured in the ultrasonic laboratory. Their ray-tracing inversion algorithm performs as well as Rayleigh-Gans (Born) inversion when the object is three wavelengths in diameter.

Also of note are some results due to Connolly (1989), who shows that the straight ray and ART iterative algorithms are modified Newton-Kantorovich and Newton-Kantorovich methods respectively. This is particularly interesting because both of these methods were originally arrived at intuitively.

4.3.4.2 Kirsch and Kress's Method

A method for approximately determining the shape of a cylinder of arbitrary cross-section has been devised by Kirsch and Kress (1983, 1987). It has been developed further by Angell *et al.* (1987) and Jones and Mao (1989). The method uses far-field scattering data for one or more incident directions. Weak *a priori* information about the shape of the scatterer is required and is readily incorporated into the formulation. It is convenient to explain the method in terms of a cylinder on whose surface Dirichlet boundary conditions are satisfied.

The first step in implementing the method is to approximately determine $\psi_s(\mathbf{x}, k)$ from the far-field data. Then, $\sigma(\mathbf{x})$ is found as a curve where the total field is zero.

To determine $\psi_s(\mathbf{x}, k)$ approximately, (3.47) is invoked in the far-field of Υ_- to give

$$\psi_{ff_s}(\hat{\mathbf{x}}, k) = \int_\sigma \frac{\partial \psi(\mathbf{x}_1, k)}{\partial \mathbf{n}} e^{jk\hat{\mathbf{x}}\cdot\mathbf{x}_1} d\sigma(\mathbf{x}_1) \tag{4.79}$$

where $\frac{\partial\psi(\mathbf{x}_1,k)}{\partial\mathbf{n}}$ can be regarded as the density of equivalent surface sources and $\sigma(\mathbf{x})$ is the initially unknown surface of the cylinder. If the equivalent surface sources could be found then $\psi_s(\mathbf{x},k)$ could easily be calculated by invoking (3.47). The difficulty, however, is that $\sigma(\mathbf{x})$ is not initially known. In order to overcome this difficulty, Kirsch and Kress (1987) approximate $\sigma(\mathbf{x})$ by an arbitrary surface which lies inside $\sigma(\mathbf{x})$. The only *a priori* knowledge required is the minimum interior envelope of $\sigma(\mathbf{x})$. This allows (4.79) to be written as a Fredholm integral equation of the first kind:

$$\psi_{ff_s}(\hat{\mathbf{x}},k)=\int_\Gamma \frac{\partial\psi(\mathbf{x}_1,k)}{\partial\mathbf{n}}e^{jk\hat{\mathbf{x}}\cdot\mathbf{x}_1}d\Gamma(\mathbf{x}_1) \tag{4.80}$$

where $\Gamma(\mathbf{x})$ is the aforesaid arbitrary surface lying inside $\sigma(\mathbf{x})$. A solution to (4.80) for the surface sources is found by invoking the algebraic expansion technique for solving integral equations explained in §2.2.8. Because the equation tends to be severely ill-posed (Kirsch and Kress, 1987), Tikhonov regularisation is invoked, so that the resultant solution is obtained by minimising

$$A_1(\frac{\partial\psi(\mathbf{x}_1,k)}{\partial\mathbf{n}},\psi_{ff_s},\alpha)=|\int_\Gamma \frac{\partial\psi(\mathbf{x}_1,k)}{\partial\mathbf{n}}e^{jk\hat{\mathbf{x}}\cdot\mathbf{x}_1}d\Gamma(\mathbf{x}_1)-\psi_{ff_s}(\hat{\mathbf{x}}_1,k)|^2+\ \alpha|\frac{\partial\psi(\mathbf{x}_1,k)}{\partial\mathbf{n}}|^2 \tag{4.81}$$

where α is the regularising parameter. The next step is to determine $\sigma(\mathbf{x})$ as that surface on which the total field is zero. This operation is eased when it can be assumed that $\sigma(\mathbf{x})$ is single valued, so that it can be parameterised as $\sigma=\sigma(\theta)$. By invoking (3.47) the scattered field can thus be expressed as

$$\psi_s(\mathbf{x},k)=\int_0^{2\pi}\frac{\partial\psi(\mathbf{x},k)}{\partial\mathbf{n}}\frac{e^{-jk\nu r}}{4\pi r}d\theta \tag{4.82}$$

where $r=|\sigma(\theta)-\mathbf{x}|$. After introducing the functional

$$A_2(\psi_s(\mathbf{x}_1,k))=\int_0^{(2\pi}|\psi_s(\mathbf{x}_1,k)|^2d\theta \tag{4.83}$$

Kirsch and Kress (1987) combine it with (4.81) in such a way that $\sigma(\mathbf{x})$ is the surface which minimises

$$L(\sigma)=\min_\sigma\left(A_1(\frac{\partial\psi(\mathbf{x}_1,k)}{\partial\mathbf{n}},\psi_{ff_s},\alpha)+\alpha_1A_2(\psi_s(\mathbf{x},k))\right) \tag{4.84}$$

Kirsch *et al.* (1988) present numerical reconstructions of $\sigma(\mathbf{x})$ for several cylinders on the surfaces of which Dirichlet boundary conditions are satisfied. Results for two of these are summarised in table 4.5. The surface $\Gamma(\mathbf{x})$ is a circle for both of these reconstructions. Kirsch *et al.* (1988) and Blohbaum (1989) compare Kirsch and Kress's method with Colton and Monk's method (see §4.3.3.4). They conclude that both methods perform equally well when more than 6 incident fields are invoked. However, Kirsch and Kress's method requires some interactive involvement in guessing an appropriate surface $\Gamma(\mathbf{x})$. Angell *et al.* (1987) have suggested that Kirsch and Kress's (1987) method could be extended by approximating the surface sources with a spherical wavefunction expansion. Jones and Mao (1989) have investigated such a spherical wavefunction expansion technique.

4.4 Final Remarks

The inverse methods presented in this Chapter can be grouped into those which provide reconstructions of $\chi(\mathbf{x})$ and those which provide reconstructions of $\sigma(\mathbf{x})$.

It is clear that the methods for reconstructing $\chi(\mathbf{x})$ have their own particular range of applicability. Ray tracing methods provide good results for fields at high frequencies. Preliminary results for the Herglotz wave function method presented in §4.3.3.4 indicate that it is applicable to objects which exhibit strong scattering (as defined in §2.2.3) when the object is circularily symmetric. The other methods provide accurate reconstructions for objects which exhibit very weak to weak scattering (as defined in §2.2.3). The ultimate aim in the development of these techniques is to make them suitable for reconstructing objects which exhibt strong scattering. Consequently, a motivation for the original inverse methods presented in the next Chapter (in §5.1 and §5.3) has been to develop methods which are suitable for reconstructing objects which exhibit strong scattering.

The methods which are able to reconstruct $\sigma(\mathbf{x})$ are relatively refined in comparison to those which provide reconstructions of $\chi(\mathbf{x})$. All these methods are based on the implicit techniques presented in §4.3. The results indicate that each method provides accurate reconstructions under many conditions. However, it is clear that the wider the angular ranges over which the incident and scattered fields are specified, the more stable and accurate are the reconstructions. It is also clear that the numerical difficulty and computational expense involved in reconstructing objects grows rapidly with increases in the linear dimensions (expressed in wavelengths) of the objects. The main motivation for the original technique presented in the next Chapter (in §5.2) has been to reduce the computational effort required to produce reconstructions of $\sigma(\mathbf{x})$. Another motivation has been to extend the range of objects which are tractable, in the sense that they can be reconstructed efficiently and accurately with readily implementable computational algorithms.

Chapter 5

New Solutions to the Inverse Scattering Problem

This Chapter introduces three original inverse methods developed with the idea of increasing the numerical efficiency and reconstruction accuracy of previously established methods. The algorithms presented in §5.1 and §5.2 are classified as implicit exact, whilst that presented in §5.3 is classified as explicit approximate. For computational economy all numerical solutions and results are presented in two dimensions, for the reasons given in §1.5.

In §5.1 a numerical solution to the inverse problem for the Helmholtz equation with spatially varying wavenumber is presented. The algorithm is based on a global solution (Bates, 1975; 1984) to the inverse problem which expresses the field (within the inhomogeneous part of space from which the wave motion is scattered) in basis functions that explicitly incorporate the scattering data. This has been shown to allow a significant reduction in the numerical effort required to generate an implicit exact solution (Tan, Murch and Bates, 1988; Tan, Murch and Bates, 1989). In §5.1.2 the theory underlying the algorithm is formulated for three-dimensional scatterers.

A method for determining the shape of an impenetrable obstacle from the complex amplitude and bistatic cross-section (see §2.2.10) of its far-field scattering pattern is presented in §5.2. The problem is formulated as a non-linear operator equation which is then solved by an iterative method. The use of the null-field method to solve the direct problem at each iteration leads to efficient numerical evaluation of the Fréchet derivative of the non-linear operator (Murch *et al.*, 1988b). Results of numerical trials with simulated data are presented in §5.2.4.

An extended Rayleigh-Gans (Born) approximation is presented in §5.3. The motivation here has been to improve the reconstruction accuracy of existing explicit approximate solutions to the inverse problem. Almost all previous explicit approximate solutions to the inverse problem are based on either the Rayleigh-Gans (Born) or Rytov approximations and any improvements to them could represent an important step towards attaining more widely applicable useful accurate efficient inverse solutions. In §5.3.1 the formal theory of the method is presented. In §5.3.2 it is explained how the method is implemented, while numerical results are presented in §5.3.3.

5.1 Global Solution to Scalar Inverse Scattering Problem

In this section a technique for generating numerical solutions to inverse problems involving the Helmholtz equation with spatially varying wave number is presented. The algorithm is based on the global formalism which was introduced some time ago (Bates, 1975) and was discussed further more recently (Bates, 1984). Although numerical methods have been previously devised, for iterating towards solutions to inverse scattering problems by solving successions of direct problems (Roger, 1981; Johnson and Tracy, 1983; Colton, 1984; Lesselier *et al.*, 1985; Kristensson and Vogel, 1986) (see §4.3.3), increased numerical efficiency is to be expected from this approach because it expresses the wave motion in basis functions that are uniquely related to the inverse scattering data.

5.1.1 Preliminaries

It is convenient to define $\sigma(\mathbf{x})$ to be a hypersphere of radius $r = a$ enclosing all spatial variations of $\nu(\mathbf{x})$. The inverse scattering problem can then be posed as determining $\nu(\mathbf{x})$ within the hypersphere Υ_-. Υ_{obs} is then taken as Υ_+. Because arbitrary variations of $\nu(\mathbf{x})$ within Υ_- are permissible, the generality of the approach is not reduced by treating Υ_- as a hypersphere.

The boundary conditions on $\sigma(\mathbf{x})$ can then be represented as

$$\psi(a-\epsilon,\phi,\theta,k) = \psi(a+\epsilon,\phi,\theta,k) \quad \text{and} \quad \partial_r\psi(a-\epsilon,\phi,\theta,k) = \partial_r\psi(a+\epsilon,\phi,\theta,k) \tag{5.1}$$

where

$$\epsilon \longrightarrow 0 \tag{5.2}$$

with ϵ being real and positive.

5.1.2 Formal solution

Consider a set $\{\psi_i^{p,l}(r,\phi,\theta,k); p = -l,-l+1,...,l; l = 0,1,...\}$ of incident fields, the p,l^{th} being chosen to be

$$\psi_{inc}^{p,l}(r,\phi,\theta,k) = j_l(kr)Y_l^p(\phi,\theta) \quad \text{for} \tag{5.3}$$

where $j_l(\cdot)$ denotes the spherical Bessel function of the first kind of order l, and $Y_l^p(\cdot,\cdot)$ is the p,l^{th} spherical harmonic regular throughout $0 \le \theta \le \pi$ and $0 \le \phi \le 2\pi$ (Morse and Feshbach, 1953, Chapter 10). The p,l^{th} scattered field $\psi_s^{p,l}$ can then be written as

$$\psi_s^{p,l}(r,\phi,\theta,k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{m,n}^{p,l}(k)h_n^{(2)}(kr)Y_n^m(\phi,\theta) \quad \text{for} \quad r > a \tag{5.4}$$

where $A_{m,n}^{p,l}(k)$ is the m,n^{th} expansion coefficient for the p,l^{th} wavefunction $\psi^{p,l}$, and $h_n^{(2)}(\cdot)$ denotes the spherical Hankel function of the second kind of order n. The $A_{m,n}^{p,l}(k)$ constitute the data for the inverse scattering problem because, as stated in §5.1.1, the wave function is known throughout Υ_+

An appropriate representation of $\psi^{p,l}$, which is the total wave function throughout Υ_-, is

$$\psi^{p,l}(r,\phi,\theta,k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \psi^{p,l}_{m,n}(r,k) Y^m_n(\phi,\theta) \quad \text{for} \quad r < a \tag{5.5}$$

where the functional forms of $\psi^{p,l}_{m,n}(r,k)$ are, as yet, unspecified. The impedances can then be defined as

$$Z^{p,l}_{m,n}(k) = -\frac{\partial_r \psi^{p,l}_{m,n}(a,k)}{\psi^{p,l}_{m,n}(a,k)} \tag{5.6}$$

Reference to (5.1), (5.3) and (5.4) then indicates that

$$Z^{p,l}_{m,n}(k) = -k \frac{j_n'(ka)\delta_{p,m}\delta_{l,n} + A^{p,l}_{m,n} h_n^{(2)'}(ka)}{j_n(ka)\delta_{p,m}\delta_{l,n} + A^{p,l}_{m,n} h_n^{(2)}(ka)} \tag{5.7}$$

where m and n are integers. Note that

$$Z^{p,l}_{m,n}(k) = -k \frac{h_n^{(2)'}(ka)}{h_n^{(2)}(ka)} \quad \text{for} \quad m \neq p \quad \text{and} \quad n \neq l \tag{5.8}$$

Since the $\psi^{p,l}_{inc}(r,\phi,\theta,k)$ are all of unit amplitude, in the sense that there is no multiplier of the standard eigenfunctions on the right hand side of (5.3), it is appropriate to think of the $Z^{p,l}_{p,l}$ and the $A^{p,l}_{m\neq p,n\neq l}(k)$ as the normalised inverse scattering data.

When (5.5) is substituted into (2.3), and the orthogonality of the $Y^m_n(\phi,\theta)$ is invoked, the resulting second order differential equations (with respect to r) seem most readily manipulable if the $\psi_{m,n}$ are defined as

$$\psi^{p,l}_{m,n}(r,k) = \sum_{\mu=1}^{\infty} B^{p,l}_{\mu,m,n}(k) j_n(\alpha^{p,l}_{\mu,m,n} r) \tag{5.9}$$

where $B^{p,l}_{\mu,m,n}(k)$ is the μ, m, n^{th} expansion coefficient for the p, l^{th} wave function. On requiring that each $\alpha^{p,l}_{\mu,m,n} = \alpha^{p,l}_{\mu,m,n}(k)$ satisfy the eigenvalue relation

$$\alpha^{p,l}_{\mu,m,n} j_n'(\alpha^{p,l}_{\mu,m,n}) = Z^{p,l}_{m,n}(k) j_n(\alpha^{p,l}_{\mu,m,n}) \tag{5.10}$$

it is seen that (5.5) automatically accommodates the normalised inverse scattering data. In fact, as has been remarked (although in contexts restricted to two space dimensions (Bates, 1975; Bates, 1984)) the $\alpha^{p,l}_{\mu,m,n}(k)$ are uniquely related to the $Z^{p,l}_{m,n}(k)$. This is because, by an obvious extension of the theory of Dini series (Watson, 1966), $\{j_n(\alpha^{p,l}_{\mu,m,n} r); \mu = 1,2,...\}$ is found to be an orthogonal set (with weight function r^2) throughout $0 \leq r \leq a$, for any quartet of integers p, l, m and n, with l and n non-negative and $|p| \leq l$ and $|m| \leq n$. Consequently, to recover ν for $r < a$, the $B^{p,l}_{\mu,m,n}(k)$ must be inferred either explicitly or implicitly, from the $Z^{p,l}_{m,n}(k)$ and $A^{p,l}_{\mu,m,n}(k)$.

On substituting (5.5) and (5.9) into (2.3), and invoking Bessel's equation (Watson, 1966) to eliminate the term $r^{-2}\partial_r(r^2 \partial_r j_n(\alpha^{p,l}_{\mu,m,n} r))$, it transpires that

$$\sum_{\tilde{n}=0}^{\infty} \sum_{\tilde{m}=-\tilde{n}}^{\tilde{n}} \sum_{\tilde{\mu}=1}^{\infty} \left[I^{\tilde{\mu},\tilde{m},\tilde{n},p,l}_{\mu,m,n} - \delta_{\mu,\tilde{\mu}} (\alpha^{p,l}_{\mu,m,n})^2 N^{p,l}_{\mu,m,n} \right] B^{p,l}_{\tilde{\mu},\tilde{m},\tilde{n}}(k) = 0 \quad \text{for} \quad r < a \tag{5.11}$$

where

$$I^{\tilde{\mu},\tilde{m},\tilde{n},p,l}_{\mu,m,n} = \int_0^a \int_0^{2\pi} \int_0^{\pi} k^2\nu^2(r,\phi,\theta) j_n(\alpha^{p,l}_{\mu,m,n} r) j_{\tilde{n}}(\alpha^{p,l}_{\tilde{\mu},\tilde{m},\tilde{n}} r) \\ Y^m_n(\phi,\theta) Y^{\tilde{m}*}_{\tilde{n}}(\phi,\theta) r^2 \sin\theta d\theta d\phi dr \tag{5.12}$$

and the normalisation constant $N^{p,l}_{\mu,m,n}(k)$ is defined by

$$N^{p,l}_{\mu,m,n}(k) = \int_0^a \int_0^{2\pi} \int_0^{\pi} \left[k j_n(\alpha^{p,l}_{\mu,m,n} r)\right]^2 Y^m_n(\phi,\theta) Y^{m*}_n(\phi,\theta) r^2 \sin\theta d\theta d\phi dr \tag{5.13}$$

where an asterisk denotes complex conjugation. (5.11) represents a triply infinite system of linear algebraic equations for the $B^{p,l}_{\mu,m,n}(k)$. A non-trivial solution exists only if the determinant of the coefficients of the $B^{p,l}_{\mu,m,n}(k)$ vanishes. No known direct means of utilising (5.11) to determine ν, nor any direct way of evaluating this "inverse scattering determinant", has been found. The following Section, demonstrates, however, that iterative manipulation of the equations derived in this Section permits the form of $\nu(r,\phi,\theta)$ to be estimated.

An intriguing aspect of the above mentioned determinant is that it suggests it may be possible to recover $\nu^2(\rho,\phi,\theta)$ for a single pair of integers p,l provided the normalised data are given for (effectively) all k. Alternatively, it would seem that a single value of k may be sufficient, if data are available for (effectively) all pairs of integers p,l. These remarks are not offered with any rigorous intent, but they are worth making because they accord with results obtained for wave motions that do not exhibit refraction (Chadan and Sabatier, 1977).

When the refractive index is spherically symmetric, so that it depends only upon the radial co-ordinate r, one would expect to be able to recover $\nu^2 = \nu^2(r)$ from data depending upon a single parameter. Inspection of (5.4) indicates that $A^{p,l}_{m,n}(k) = 0$ unless $m = p$ and $n = l$. So, it is expected that the $A^{p,l}_{p,l}(k)$ represent sufficient data either for a single pair of integers p,l and a range of k or a single value k and a range of pairs of integers p,l. This accords with what has been established for the non-relativistic quantum mechanical inverse scattering problem(Chadan and Sabatier, 1977). Expressions which are to be evaluated numerically necessarily contain finite numbers of parameters. This implies that to be able to estimate $\nu^2(r)$ usefully given $A^{p,l}_{m,n}(k)$ at a finite number, say $\hat{\mu}$, of values k_j of k, for a single pair of integers p,l. It is also expected that to reconstruct $\nu^2(r)$ to the same order of accuracy from $A^{p,l}_{m,n}(k)$ given for only one value of k, a single integer p, and about the same number $\hat{\mu}$ of integers l. These expectations are borne out by the results presented in §5.1.4.

Since a non-spherically symmetric refractive index requires many more parameters to characterise it to the same accuracy than one that is spherically symmetric, more data should be needed to reconstruct it usefully. This does not necessarily imply, however, that more discrete values of k are required or more pairs of integers p,l. The point is that for a single pair of integers p,l and one value of k, there are many expansion coefficients $A^{p,l}_{m,n}(k)$. So, provided the $A^{p,l}_{m,n}(k_j)$ are given for J values of k_j for a single pair of integers p,l and for all pairs of integers m,n for which $|A^{p,l}_{m,n}|$ is significant, one expects to be able to recover a useful estimate of $\nu^2(r,\phi,\theta)$. This means that one must supplement the inverse scattering determinant with equations relating the $B^{p,l}_{\mu,m,n}(k)$ to the $A^{p,l}_{m,n}(k)$ for $m \neq p$ and $n \neq l$.

5.1.3 Algorithmic Implementation of the Formal Solution

All the conceptual and analytical difficulties associated with this approach (and all other exact approaches, for that matter) to the inverse scattering problem are present in two space dimensions, although the required computational effort is of course much reduced by comparison with three dimensions §1.5. So, as already intimated in §1.5, "computational energy" is conserved by operating on the two-dimensional versions of the equations developed in §5.2. The three-dimensional forms of (5.3) through (5.7) and (5.9) through (5.13) are replaced by their two-dimensional counterparts. These are

$$\psi^l_{inc}(\rho,\phi,k) = J_l(k\rho)e^{jl\phi} \tag{5.14}$$

$$\psi^l_s(\rho,\phi,k) = \sum_{m=0}^{\infty} A^l_m(k)H^{(2)}_m(k\rho)e^{jm\phi} \quad \text{for} \quad \rho > a \tag{5.15}$$

$$\psi^l(\rho,\phi,k) = \sum_{m=-\infty}^{\infty} \psi^l_m(\rho,k)e^{jm\phi} \quad \text{for} \quad \rho \le a \tag{5.16}$$

$$Z^l_m(k) = -\frac{\partial_\rho \psi^l_m(a,k)}{\psi^l_m(a,k)} = -k\frac{J'_m(ka)\delta_{l,m} + A^l_m H^{(2)'}_m(ka)}{J_m(ka)\delta_{l,m} + A^l_m H^{(2)}_m(ka)} \tag{5.17}$$

$$\psi^l_m(\rho,k) = \sum_{\mu=1}^{\infty} B^l_{\mu,m}(k)J_m(\alpha^l_{\mu,m}\rho) \quad \text{for} \quad \rho \le a \tag{5.18}$$

$$\alpha^l_{\mu,m}J'_{\mu,m}(\alpha^l_{\mu,m}) = Z^l_m(k)J_m(\alpha^l_{\mu,m}) \tag{5.19}$$

$$\sum_{\tilde{m}=-\infty}^{\infty}\sum_{\tilde{\mu}=1}^{\infty}\left[I^{\tilde{\mu},\tilde{m},l}_{\mu,m} - \delta_{\mu,\tilde{\mu}}(\alpha^l_{\mu,m})^2 N^l_{\mu,m}\right] B^l_{\tilde{\mu},\tilde{m}}(k) = 0 \tag{5.20}$$

$$I^{\tilde{\mu},\tilde{m},l}_{\mu,m} = \int_0^a \int_0^{2\pi} k^2\nu^2(\rho,\phi)J_m(\alpha^l_{\mu,m}\rho)J_{\tilde{m}}(\alpha^l_{\tilde{\mu},\tilde{m}}\rho)e^{jm\phi}e^{-j\tilde{m}\phi}\rho d\phi d\rho \tag{5.21}$$

$$N^l_{\mu,m} = \int_0^a \int_0^{2\pi} \left[kJ_m(\alpha^l_{\mu,m}\rho)\right]^2 \rho d\phi d\rho \tag{5.22}$$

Note that, because there is one less dimension, the quartet p, l, m, n of integers introduced in §5.1.2 reduces here to the pair l, m.

The $Z^l_m(k)$ and the $A^l_m(k)$ constitute the normalised data for the two-dimensional inverse scattering problem. Because a direct method of evaluating the $I^{\tilde{\mu},\tilde{m}}_{\mu,m}$ has not been found, an iterative scheme has been devised and is introduced below.

The first step in the solution is to solve the eigenvalue equation (5.19) to obtain $\alpha^l_{\mu,m}$ for all the needed values of l, μ and m. This is accomplished by a Newton root finding algorithm. The eigenvalues are spaced by close to π for large μ, so the initial estimate of $\alpha^l_{\mu+1,m}$ is taken to be $\alpha^l_{\mu,m} + \pi$. In this way the desired number of eigenvalues can be readily determined numerically.

The unknown $\nu^2(\rho,\phi)$ is expressed in terms of an appropriate set of basis functions. The representation chosen is

$$\nu^2(\rho,\phi) = \sum_{n=1}^{N}\sum_{p=-P}^{P} C_{n,p}\text{rect}_n(\rho)e^{jp\phi} \tag{5.23}$$

where the $C_{n,p}$ are expansion coefficients, with $C_{n,p} = C^*_{n,-p}$ to ensure ν is real, and

$$\text{rect}_n(\rho) = \begin{cases} 1 & \text{for } \frac{(n-1)a}{N} < \rho < \frac{na}{N} \\ 0 & \text{otherwise} \end{cases} \tag{5.24}$$

and N and P are finite positive integers. The finiteness of N and P emphasises that the algorithm can provide numerical solutions. The algorithm cannot of course be implemented for infinite numbers of basis functions (or for continuous functions either because they have to be quantised in amplitude and spatially sampled, both of which are to be understood in what follows). Finite sets of the integers l, m and p must also be chosen.

Once the $\alpha^l_{\mu,m}$ have been determined, it is expedient to compute the quantities

$$\mathcal{I}^{l,\mu,m,\tilde{\mu},\tilde{m}}_n = \int_0^a k^2 \text{rect}_n(\rho) J_m(\alpha^l_{\mu,m}\rho) J_{\tilde{m}}(\alpha^l_{\tilde{\mu},\tilde{m}}\rho)\rho d\rho \tag{5.25}$$

for later use. An adaptive Simpson algorithm is employed for these computations (Conte and de Boor, 1980).

Because the method is iterative, it needs an initial estimate of ν, characterised by a set $\{C^{(0)}_{n,p}\}$ of initial expansion coefficients, in order to start it. The remainder of this section indicates how sets of expansion coefficients, which are improved in the sense that they increasingly approach the set $\{C_{n,p}\}$ for the true refractive index, are obtained at each iteration. A superscript (i) attached to a quantity identifies the estimate generated at the i^{th} iteration.

When (5.23) is substituted into (5.21), and (5.14) through (5.22) and (5.25) are invoked, it is seen that

$$\sum_{\tilde{m}=-M}^{M} \sum_{\tilde{\mu}=1}^{\hat{\mu}} \left[\sum_{n=1}^{\infty} C_{n,m-\tilde{m}} \mathcal{I}^{l,\mu,m,\tilde{\mu},\tilde{m}}_n - \delta_{\mu,\tilde{\mu}} (\alpha^l_{\mu,m})^2 N^l_{\mu,m} \right] B^l_{\tilde{\mu},\tilde{m}} = 0 \tag{5.26}$$

where $\mu \in \{1, 2, ..., \hat{\mu}\}, m \in \{-M, -M+1, ..., M\}$ and $\hat{\mu}$ and M are finite positive integers. Let $l \in \{0, 1, 2, ..., L\}$ with L being a finite non-negative integer. This gives a set of homogeneous equations for the $B^l_{\mu,m}$. Equations (5.14), (5.15), (5.16) and (5.18) provide a set of non-homogeneous equations relating the given A^l_m to the $B^l_{\mu,m}$, namely

$$\sum_{\mu=1}^{\infty} B^l_{\mu,m} J_m(\alpha^l_{\mu,m} a) = A^l_m H^{(2)}_m(ka) + \delta_{l,m} J_l(ka) \tag{5.27}$$

Equations (5.26) and (5.27) need to be solved simultaneously. This is performed by combining (5.26) and (5.27) for all l as a single set of non-homogeneous equations and writing them for the i^{th} iteration as

$$\mathbf{K}(\mathbf{C}^{(i)})\mathbf{B}^{(i)} = \mathbf{A} \tag{5.28}$$

where the components of the vectors $\mathbf{C}^{(i)}, \mathbf{B}^{(i)}$ and $\mathbf{A}^{(i)}$ are the $C^{(i)}_{n,p}, B^{l(i)}_{\mu,m}$ and A^l_m respectively. The elements of the matrix $\mathbf{K}$ are straightforwardly found from inspection of (5.26) and (5.27). At each iteration, (5.28) is solved in a least squares sense and the residue $\mathbf{F}^{(i)}$ of this least squares solution is computed as

$$\mathbf{F}^{(i)}(\mathbf{C}^{(i)}) = \mathbf{A} - \mathbf{K}(\mathbf{C}^{(i)})\mathbf{B}^{(i)} \tag{5.29}$$

The functional dependence on $\mathbf{C}^{(i)}$ emphasises that $\mathbf{K}$ is evaluated using the current estimate of the refractive index. Next the Fréchet derivative (see §4.3.2.1) of $\mathbf{F}$ is formed by calculating central differences.

An update to $\mathbf{C}^{(i)}$ is found by solving (in a least squares sense)

$$\mathbf{F}'^{(i)}(\mathbf{C}^{(i)})\delta\vec{\mathbf{C}}^{(i)} = -\mathbf{F}^{(i)}(\mathbf{C}^{(i)}) \tag{5.30}$$

where $\mathbf{F}'^{(i)}$ is the Fréchet derivative at iteration (i) and

$$\mathbf{C}^{(i+1)} = \mathbf{C}^{(i)} + \delta\mathbf{C}^{(i)} \tag{5.31}$$

The iterations are terminated when

$$||\mathbf{C}^{(i+1)} - \mathbf{C}^{(i)}||_\infty \leq \epsilon \tag{5.32}$$

where ϵ is a pre-set level and the infinity norm is defined in §2.2.9. $\hat{i}$ denotes the number of iterations which have to be performed before the inequality (5.32) is satisfied. The faithfulness with which $\nu^2(\rho)$ has then been reconstructed is characterised by

$$\Delta = ||\mathbf{C} - \mathbf{C}^{(i)}||_\infty \tag{5.33}$$

where $\mathbf{C}$ is the vector whose components are the actual $C_{n,p}$.

Equations (5.30) and (5.31) represent the Newton-Kantorovich method (see §4.3.2.1) of solving the non-linear operator equation

$$\mathbf{F}(\mathbf{C}^{(i)}) = \mathbf{0} \tag{5.34}$$

which is satisfied at the true solution but not elsewhere (locally at least).

The above represents a numerical scheme which is equivalent to forcing the determinant of the coefficients of $B^l_{\mu,m}$ in (5.20) to vanish while accommodating all of the normalised inverse scattering data.

5.1.4 Numerical Examples

To illustrate the performance of the algorithm several numerical examples are presented here. In order to obtain accurate computer-generated inverse scattering data, the scattering from two-dimensional objects of the type depicted in fig. 5.1 is computed. The interior of each object is subdivided by circular boundaries, which can of course exceed in number the four shown in fig. 5.1. Each region between such boundaries is of a constant refractive index. Each refractive index is real and non-negative but is otherwise arbitrary. The circular boundary of radius a is called the exterior circular boundary and corresponds to $\sigma(\mathbf{x})$.

The wavefunction in each region of a body (typified by fig. 5.1) was written as a finite eigenfunction expansion (involving Bessel and trigonometric functions) satisfying the Helmholtz equation in the region. The expansion coefficients were evaluated by matching the eigenfunction expansions across the circular boundaries, as indicated in §3.5. The technique is easy enough to transcribe into a computational algorithm.

The examples presented in §5.1.4.1 are included to demonstrate that the algorithm does exhibit numerical convergence, even though the series representations for ν and ψ incorporated into the algorithm are necessarily approximate (because they are unavoidably truncated). The results reported in §5.1.4.2 confirm that the algorithm can be successfully invoked for asymmetrical scatterers.

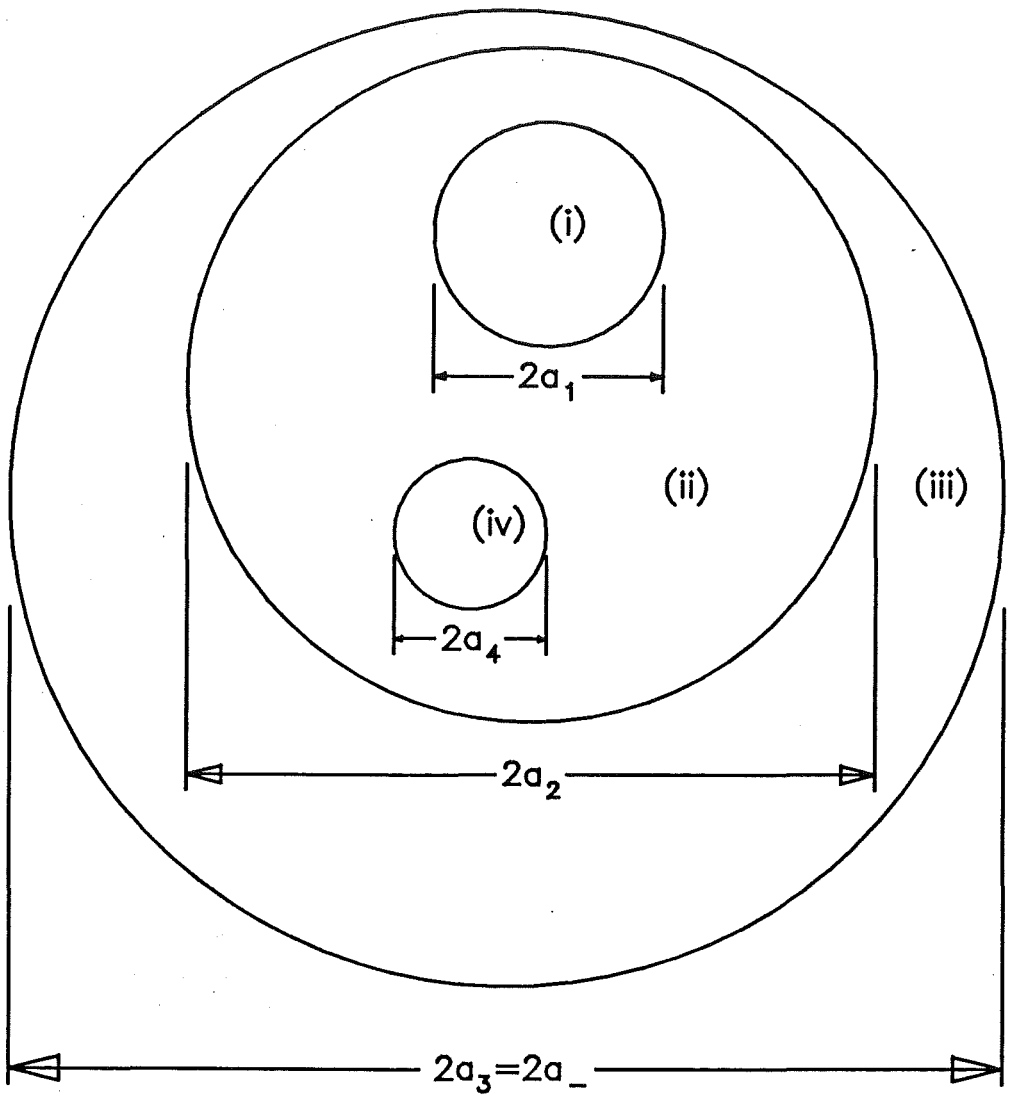


Figure 5.1: Piecewise two-dimensional object subdivided by circular boundaries. The particular object depicted here has four boundaries. The labelling of the four regions is explained in the text.

	a_q / ν_q							
Q \ q	1	2	3	4	5	6	7	8
1	1.0 11.0							
2	.5 10.0	1.0 11.0						
3	.33 10.0	.67 10.5	1.0 11.0					
4	.25 10.0	.5 10.33	.75 10.67	1.0 11.0				
5	.2 10.0	.4 10.3	.6 10.5	.8 10.8	1.0 11.0			
6	.17 10.0	.34 10.2	.5 10.4	.67 10.6	.84 10.8	1.0 11.0		
7	.14 10.0	.29 10.17	.43 10.33	.57 10.5	.71 10.67	.86 10.83	1.0 11.0	
8	.13 10.0	.25 10.14	.38 10.29	.5 10.43	.63 10.57	.75 10.71	.88 10.83	1.0 11.0

Table 5.1: Characteristics of circularly symmetric objects (having $a_- = a$) for which results are presented in table 5.2. Note that the upper and lower entries in each box are values of, respectively, a_q and ν_q.

5.1.4.1 Circularly Symmetric Examples

The required computational effort is greatly reduced when $\nu(\rho, \phi) = \nu(\rho)$. It then transpires that

$$\psi_m^l(\rho, \phi, k) = \psi_l^l(\rho, k) e^{jl\phi} \tag{5.35}$$

implying for each l, that $A_m^l(k) \equiv 0$ for $m \neq l$. Furthermore, $C_{n,p} \equiv 0$ for $p \neq 0$.

To force ν to be circularly symmetric, all the circular boundaries, which subdivide the object, are chosen to be concentric. In general there are Q boundaries, with the radius of the q^{th} being a_q and $a_Q = a_-$. The refractive index between the $(q-1)^{th}$ and q^{th} boundaries is denoted by ν_q. Reference to fig. 5.1 indicates how this can be arranged for $Q = 3$. The refractive index of region (iv) is set equal to that of region (ii), so that region (iv) effectively disappears. The boundaries of regions (i) and (ii) are then made concentric with the circle of radius $a_3 = a_-$. The constant refractive indices of regions (i) through (iii) are denoted by ν_1, ν_2 and ν_3 respectively.

Results have been obtained for the objects characterised by the values of Q, a_q and ν_q listed in table 5.1. The form of the object when $Q = 4$ is illustrated in fig. 5.2.

The performance of the algorithm can be gauged from table 5.2, which lists results for a single value of $\hat{\mu}$ but a selection of values of N. In each case $L = (N-1)$ and the initial estimate for $\nu^2(\rho)$ is chosen by setting $C_{n,0}^{(0)} = 10.0$ for $n = 1, 2, \cdots, N$. Note that numerical convergence is manifest in all cases apart from that corresponding to $Q = 1, N = 2$. Even for this latter case, however, it was found that convergence was achieved ($\hat{i} = 4$ and $\Delta = 0.002$) when L was increased to $(N+1)$. For each entry in table 5.2, the reconstruction of $\nu(\rho)$ was visually indistinguishable from the original.

Results are also presented to indicate how this algorithm performs when the data are contaminated by noise. To this end, results are presented for the objects characterised by the values of $Q = N$ in table 5.1 with $Q = 1, 2, \cdots, 8$. Again for each $Q, N = Q, \hat{\mu} = 6$ and the initial estimate for $\nu^2(\rho)$ is defined by setting $C_{n,0}^{(0)} = 10.0$ for $n = 1, 2, \cdots, Q$. The performance of the algorithm can be gauged from table 5.3, which lists results for

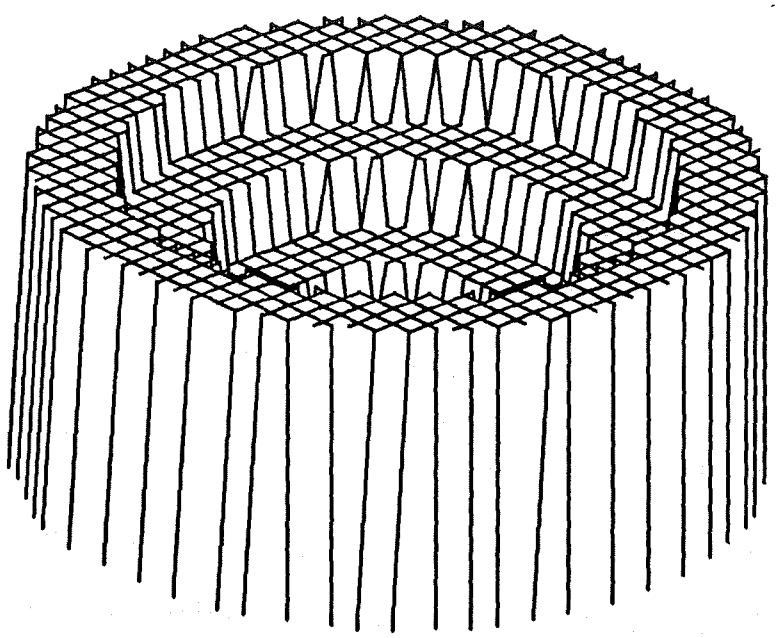


Figure 5.2: Relief map of the circularly symmetric object, whose characteristics are listed in table 5.1, for which $Q = 4$. Note that the refractive index decreases towards the centre of the object.

Q \ N	1	2	3	4	5	6	7	8
1	2 2×10^{-5}	did not converge	5 1×10^{-2}	4 1×10^{-4}	3 1×10^{-4}	4 1×10^{-3}	4 1×10^{-3}	5 1×10^{-2}
2		2 2×10^{-3}		3 1×10^{-3}		6 2×10^{-2}		6 1×10^{-2}
3			3 1×10^{-3}			5 2×10^{-2}		
4				4 1×10^{-2}				7 3×10^{-2}
5					2 5×10^{-4}			
6						9 5×10^{-2}		
7							11 5×10^{-2}	
8								11 5×10^{-2}

Table 5.2: Convergence results, obtained with the algorithm introduced in §5.1.3, for the objects listed in table 5.1 for a selection of values of N, with $ka = 1.0, \epsilon = 0.005, \hat{\mu} = 6$ and $L = (N - 1)$. Note that the upper and lower entries in each box are values of, respectively, $\hat{i}$ and Δ.

noiseless scattering data and also scattering data contaminated with additive gaussian noise having levels (as defined in §2.2.11) of 10 % and 30 % respectively. For the noiseless cases $L = N - 1$. For the additive noise cases results are shown for $L = N - 1$ and $L = N + 1$. Note that numerical convergence is manifest in all cases.

5.1.4.2 Asymmetric Examples

The goal in this section is to demonstrate that the algorithm introduced in §5.1.3 can be effective for asymmetric objects.

The simplest type of asymmetric problem is one for which the scatterer is actually symmetric but the solution is formulated on the assumption that the scatterer may be asymmetric. The scattering from a circular body of constant refractive index ($\nu = 10.5$) was computed and then submitted the algorithm to two tests (identified below by T=1 and 2). The parameters pertaining to each of the tests and the results obtained by applying the algorithm are listed in table 5.4 and its caption. The small values of Δ confirm that the algorithm converges to the correct value of ν.

The algorithm has also been applied to the explicitly asymmetric body shown in fig. 5.3. The caption to this figure lists the values to which parameters were set for the two tests (identified below by T=3 and 4) of the algorithm discussed below. The form of the object when T=3 is illustrated in fig. 5.4.

The simplest initial estimate of ν is a constant. Furthermore, the algorithm is likely to be most useful if it can converge starting from such an estimate, which is chosen arbitrarily to be 10.0. For each test of the algorithm (when applied to the scatterer shown in fig. 5.3), $L = 6, M = 3$ and $\hat{\mu} = 5$. Table 5.4 lists the values of the remaining parameters adopted for the two tests and the convergence results obtained by applying the algorithm. The performance of the algorithm can perhaps be more readily appreciated by explicitly stating the formulas it generates for $\nu(\rho; \phi)$:

	Symmetric Examples				
	noiseless	10 %		30 %	
Q \ L	L=N-1	L=N-1	L=N+1	L=N-1	L=N+1
1	2 2×10^{-5}	2 8×10^{-2}	2 3×10^{-2}	2 2×10^{-1}	2 8×10^{-2}
2	2 2×10^{-3}	2 6×10^{-1}	3 2×10^{-1}	4 5×10^{-1}	5 3×10^{-1}
3	3 1×10^{-3}	4 4×10^{-1}	5 4×10^{-1}	4 2×10^{0}	5 1×10^{0}
4	4 1×10^{-2}	4 7×10^{-1}	4 3×10^{-1}	5 1×10^{0}	6 7×10^{-1}
5	2 5×10^{-4}	3 1×10^{-1}	3 8×10^{-2}	5 1.8×10^{0}	5 5×10^{-1}
6	9 5×10^{-2}	10 5×10^{-1}	10 2×10^{-1}	11 1.2×10^{0}	11 5×10^{-1}
7	11 5×10^{-2}	11 3×10^{-1}	11 1.4×10^{0}	11 8×10^{-1}	11 1.4×10^{0}
8	11 5×10^{-2}	11 8×10^{-1}	12 2×10^{0}	12 2×10^{0}	12 4×10^{0}

Table 5.3: Convergence results, obtained with the algorithm introduced in §5.1.3, for the objects listed in table 5.1 for $Q = N$ when the data are contaminated by noise. Note that the upper and lower entries in each box are values of, respectively, $\hat{i}$ and Δ.

			$C_{n,p}^{(0)}$								
		P	-3	-2	-1	0	1	2	3		
T	n									$\hat{i}$	Δ
1	1		0.4	0.4	0.4	10.0	0.4	0.4	0.4	11	1.0×10^{-4}
2	1 2				0.4 0.4	10.0 10.0	0.4 0.4			4	1.0×10^{-4}

Table 5.4: Initial estimates of $C_{n,p}$ and results for symmetric body ($ka = 1$ and $\nu = 10.5$) assumed initially to be asymmetric. The integer T identifies each of the two tests ($L = 6, \mathrm{M} = 3, \hat{\mu} = 5, \epsilon = 0.005$ for both) of the algorithm introduced in §5.1.3 and further described in §5.1.4.2.

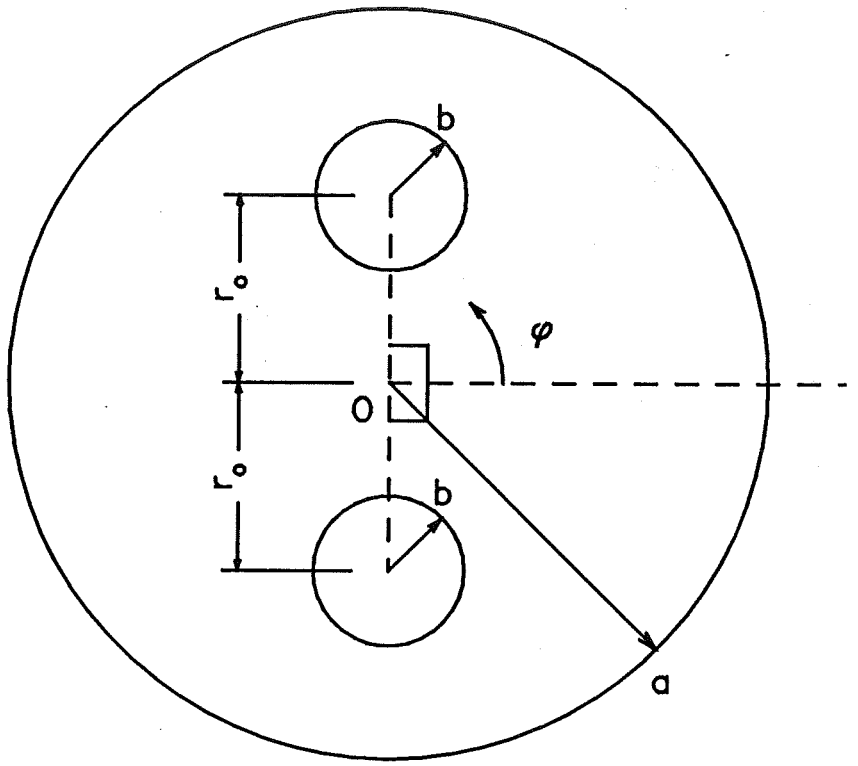


Figure 5.3: Asymmetric scatterer described in §5.1.4.2. Refractive indices of upper and lower circular regions of radius b are respectively 10.3 and 10.7. Refractive index of remainder of circular region of radius a (which is centered at O) is 10.5.

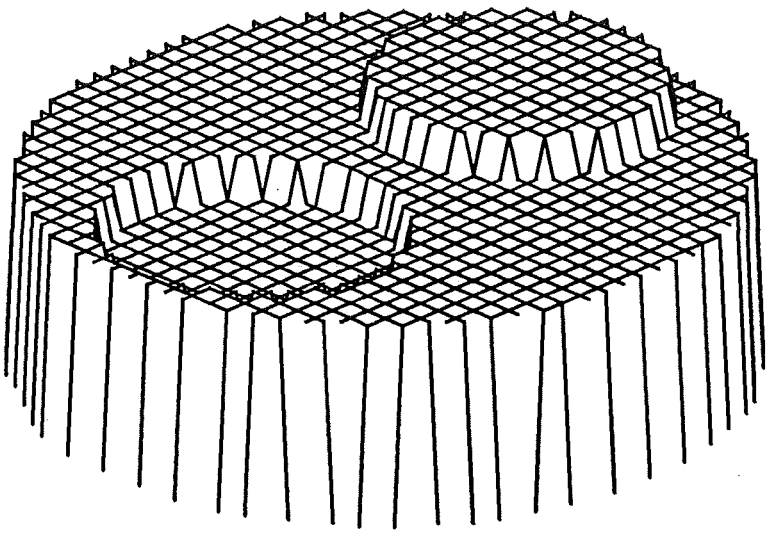


Figure 5.4: Relief map of the asymmetric scatterer for the test T=3 described in §5.1.4.2.

T	b	r_0	N	P	$\hat{i}$	Δ
3	0.45	0.5	1	3	4	.0010
4	0.2	0.75	2	1	3	.0057

Table 5.5: Parameters and convergence results for the two tests (T=3 and 4) of the algorithm introduced in §5.1.3 and further described in §5.1.4.2, applied to the scatterer shown in fig. 5.3, with $ka = 1.0$.

$$\begin{aligned} \mathbf{T=3:}\quad \nu(\rho;\phi) = \; & 10.5 \; +.0001\cos\phi - .0010\cos 2\phi + .0001\cos 3\phi \\ & +.110\sin\phi - .0010\sin 2\phi + .065\sin 3\phi \end{aligned} \tag{5.36}$$

$$\begin{aligned} \mathbf{T=4:}\quad \nu(\rho;\phi) &= 10.5 + .0001\cos\phi + .0057\sin\phi, \quad 0 < \rho \leq a/2 \\ &= 10.5 - .0016\cos\phi + .0251\sin\phi, \quad a/2 < \rho \leq a \end{aligned} \tag{5.37}$$

Both tests predict the average value of ν to within the error levels Δ listed in table 5.5. Both tests locate the maximum perturbation of $\nu(\rho;\phi)$ from its average value close to $\phi = \pm\pi/2$, which of course accords with fig. 5.3. The accuracy of this location can be gauged by comparing the amplitudes of the $\sin\phi$ and $\sin 3\phi$ terms as compared with those of, respectively, the $\cos\phi$ and $\cos 3\phi$ terms. The smallness of the amplitudes of the $\sin 2\phi$ and $\cos 2\phi$ terms for T=3, as compared with the amplitudes of the $\sin\phi$ and $\sin 3\phi$ terms, also emphasises the accuracy of the estimate provided by the algorithm. The final test (T=4) successfully locates the perturbation of the refractive index in the outer part of the scatterer, as can be seen by comparing the amplitudes of the $\sin\phi$ terms for $\rho \leq a/2$ and $\rho > a/2$ respectively.

Fig. 5.5 provides visual comparisons of the angular variations of the original (solid curves) and reconstructed (dashed curves) refractive indices, at one radius for T=3 and for two radii for T=4. The considerable differences between the dashed and solid curves in the upper plot are due to the reconstruction of $\nu(\rho;\phi)$ containing no angular Fourier coefficients of order higher than 3. The dashed curve does, however, oscillate about the solid curve and follows the general trend. The same can be said for the middle plot, which also displays the version of its solid curve obtained by discarding all of its Fourier coefficients of order higher than those invoked in the version of $\nu(\rho;\phi)$ for T=4. The closeness of the dashed and dotted curves is gratifying, as are the small differences between the solid and dashed curves in the lower plot. The middle and lower plots emphasize the faithfulness with which the algorithm can recover Fourier coefficients of $\nu(\rho;\phi)$. A relief plot of the reconstruction for T=3 is shown in fig. 5.6.

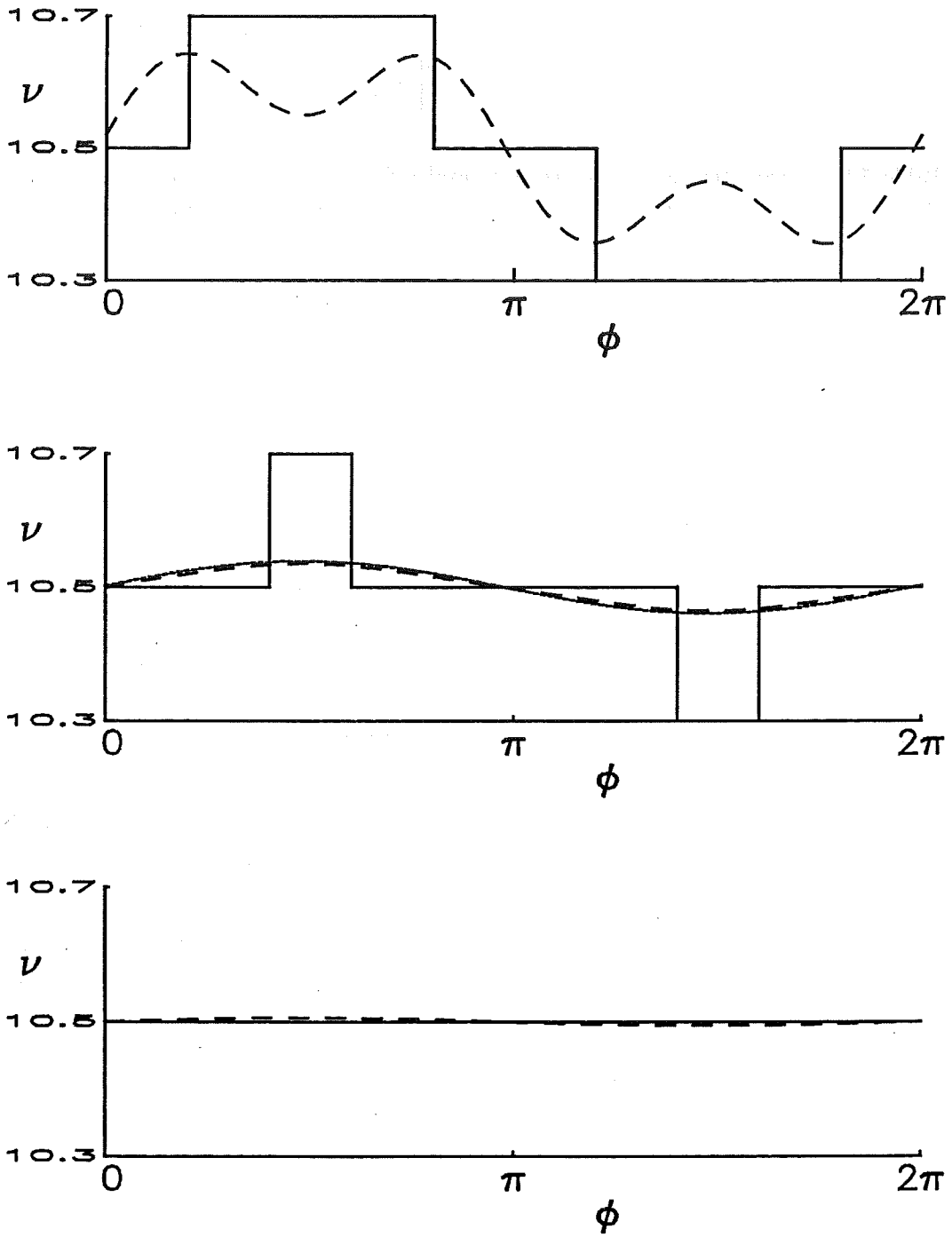


Figure 5.5: Comparison of angular variation of original (solid curves) and reconstructed (dashed curves) refractive indices for the asymmetric scatterer described in §5.1.4.2. Upper plot : T=3, $\rho = 0.5$. Middle plot : T=4, $\rho = 0.75$ (dotted curve is version of solid curve obtained by discarding all Fourier coefficients of order higher than 1). Lower plot : T=4, $\rho = 0.25$.

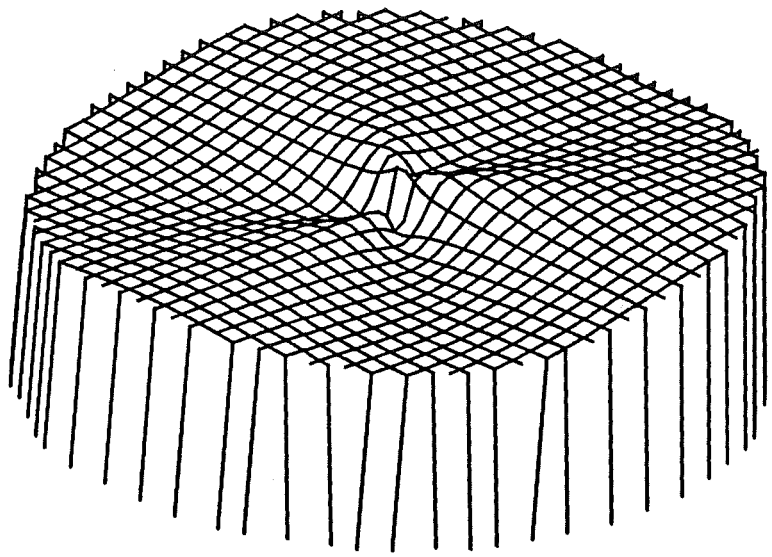


Figure 5.6: Relief map of the reconstructed asymmetric scatterer for the test T=3 described in §5.1.4.2.

5.2 Inverse Scattering for an Exterior Helmholtz Problem

In this section, the Newton-Kantorovich method is applied to an inverse problem whose associated direct problem is an exterior Helmholtz equation. The obstacle is such that the Dirichlet or Neumann boundary condition applies on its surface, as is appropriate for the tangential electric field on a perfectly conducting obstacle, or for the velocity potential on a sound soft or hard acoustic obstacle. This problem has been considered before (Roger, 1981; Colton, 1984; Wall *et al.*, 1985; Colton and Monk, 1986; Kristensson and Vogel, 1986)(see §4.3.3) and is known to be ill-posed (Colton, 1984). What is original about the results reported here is that the formulation of the direct problem employs the null-field method (see §3.3) in a manner that enables simple, efficient, and stable numerical evaluation of the Fréchet derivative (recall the discussion of the Newton-Kantorovich method in §4.3.2.1) (Murch *et al.*, 1988b).

5.2.1 Direct Problem

The null-field equations presented in §3.3 for the solution of scattering from a cylinder with either Dirichlet or Neumann boundary are reproduced here for convenience. For the Dirichlet boundary condition they are

$$f(\theta) = \sqrt{\sigma^2(\theta) + (d\sigma/d\theta)^2}\,\frac{\partial \psi(\sigma(\theta),\theta,k)}{\partial \mathbf{n}} \tag{5.38}$$

$$\sum_{n=-\infty}^{\infty} H_{mn} f_n = U_m, \quad m,n = -\infty,\ldots,\infty, \tag{5.39}$$

where U_m is defined in (3.55),

$$H_{mn} = C_m \int_0^{2\pi} H_m^{(2)}(k\sigma(\theta)) e^{j(n-m)\theta}\, d\theta. \tag{5.40}$$

$$\psi_s \sim (1/8\pi kr)^{-1/2} e^{-j(kr-3\pi/4)} \sum_{m=-\infty}^{\infty} b_m j^m e^{jm\phi}, \tag{5.41}$$

where the b_m are given by

$$b_m = \sum_{n=-\infty}^{\infty} J_{mn} f_n, \tag{5.42}$$

and

$$J_{mn} = \int_0^{2\pi} J_m(k\sigma(\theta)) e^{j(n-m)\theta}\, d\theta. \tag{5.43}$$

The equations for the Neumann boundary condition are the same except that (5.38), (5.40), and (5.43) are replaced by

$$f(\theta) = \psi(\sigma(\theta),\theta,k)/\sigma(\theta), \tag{5.44}$$

$$H_{mn} = -C_m \int_0^{2\pi} \left[k\sigma^2(\theta) H_m^{(2)\prime}(k\sigma(\theta)) + jm\sigma'(\theta) H_m^{(2)}(k\sigma(\theta))\right] e^{j(n-m)\theta}\, d\theta, \tag{5.45}$$

$$J_{mn} = -\int_0^{2\pi} \left[k\sigma(\theta) J_m'(k\sigma^2(\theta)) + jm\sigma'(\theta) J_m(k\sigma(\theta))\right] e^{j(n-m)\theta}\, d\theta. \tag{5.46}$$

5.2.2 General Newton-Kantorovich Algorithm

The Newton-Kantorovich formulation provides a general method which can be employed in many applications (see §4.3.2.1). The form of Λ' depends, of course, on Λ, and the success of the method depends on the reliable calculation, at each iteration, of both of these quantities. For the direct problem considered here, the mapping of a particular obstacle to its scattered field is computed using the null-field equations listed in §5.2.1. It remains to describe, in the next section, a method for computing the Fréchet derivative of the direct mapping.

5.2.3 Computation of Fréchet Derivative

The operator Λ considered here is defined as a set of normalised scattering data due to a single incident wave, specified in the far-field at a number of different angles, denoted ϕ_{meas}. This mapping is conveniently expressed by the functional

$$\Lambda(\sigma) = \sum_{m=-\infty}^{\infty} j^m b_m e^{jm\phi_{meas}}, \tag{5.47}$$

where the functional dependence on $\sigma(\theta)$ is implicit in $\{b_m\}$ through (5.39), (5.40), (5.42), and (5.43). It is appropriate to introduce a representation for $\sigma(\theta)$ in terms of some expansion coefficients $\{\rho_p\}$ and basis functions $\{\varrho_p(\theta)\}$:

$$\sigma(\theta) = \sum_p \rho_p \varrho_p(\theta), \tag{5.48}$$

where, for the moment, the number of basis functions and their functional form is arbitrary. The Fréchet derivative is then a matrix operator with elements that are merely the partial derivatives of Λ with respect to each ρ_p. The element corresponding to ϕ_{meas} and ρ_p is

$$\Lambda'_p = \partial\Lambda/\partial\rho_p = \sum_{m=-\infty}^{\infty} j^m \partial b_m/\partial\rho_p e^{jm\phi_{meas}}. \tag{5.49}$$

From (5.42), the term $\partial b_m/\partial\rho_p$ is conveniently expressed as

$$\partial b_m/\partial\rho_p = \sum_{n=-\infty}^{\infty} (J'_{mnp} f_n + J_{mn} f'_{np}). \tag{5.50}$$

Note that the primes in the concise notation J'_{mnp} and f'_{np} do not denote differentiation. In fact, from (5.43),

$$\begin{aligned} J'_{mnp} &= k \int_0^{2\pi} J'_m(k\sigma(\theta)) \frac{\partial\sigma}{\partial\rho_p}(\theta) e^{j(n-m)\theta}\, d\theta \\ &= k \int_0^{2\pi} J'_m(k\sigma(\theta)) \varrho_p(\theta) e^{j(n-m)\theta}\, d\theta. \end{aligned} \tag{5.51}$$

The remaining term, in (5.50), yet to be defined, is f'_{np}. This can be computed by noting that (5.39), when differentiated and rearranged reads, for each $p = -\infty, \ldots, \infty$,

$$\sum_{n=-\infty}^{\infty} H_{mn} f'_{np} = -\sum_{n=-\infty}^{\infty} H'_{mnp} f_n, \quad m = -\infty, \ldots, \infty, \tag{5.52}$$

where H'_{mnp} is defined similarly to J'_{mnp}, that is, (5.51) with J_m replaced by $C_m H_m^{(2)}$. Note that, regardless of the value of p, the linear system of equations (5.52) involves the same linear operator as (5.39). Since only the right hand sides differ, the set $\{f_{np}\}$ can be calculated with little extra work once the inverse operator has been computed to solve (5.39).

Efforts to efficiently compute J'_{mnp} and H'_{mnp} influence the choice of $\{\varrho_p(\theta)\}$, as is demonstrated in §5.2.4.

It is worth noting that, in practical situations, accurate measurement of the phase of ψ_s may be difficult (Bates and Tan, 1985) so that a more realistic measurement is of the intensity

$$\mathcal{U}(\sigma) = \Lambda(\sigma)\Lambda^*(\sigma). \tag{5.53}$$

This alternative form of direct mapping causes no serious complication to the procedures developed above, because each component of the Fréchet derivative of $\mathcal{U}$ is easily evaluated once Λ' has been computed with the aid of (5.49):

$$\partial\mathcal{U}/\partial\rho_p = 2\mathrm{Re}(\Lambda^*\partial\Lambda/\partial\rho_p) \tag{5.54}$$

Reference to (5.46) reveals that, when the Neumann boundary condition applies, (5.51) should be replaced by

$$\begin{aligned} J'_{mnp} \;=\; & -\int_0^{2\pi} \Big\{ \Big[k(2\sigma(\theta) + jm\sigma'(\theta)) J'_m(k\sigma(\theta)) + k^2\sigma^2(\theta) J''_m(k\sigma(\theta)) \Big] \varrho_p(\theta) \\ & + jm J_m(k\sigma(\theta)) \varrho'_p(\theta) \Big\} e^{j(n-m)\theta}\, d\theta \end{aligned} \tag{5.55}$$

the double prime denoting second derivatives, with the expression for H'_{mnp} being the same except for J_m replaced by $C_m H_m^{(2)}$.

5.2.4 Numerical Implementation and Results

Computation of Λ' entails summation of the expressions in (5.49) and (5.50), and solution of the system of equations (5.52) for all p. Note that $\{J_{mn}\}$ and $\{f_n\}$ can be retrieved from the solution of the direct problem, and require no further computation. It remains to specify particular basis functions $\{\varrho_p(\theta)\}$ that ease the effort in solving the linear equations resulting from the Newton-Kantorovich method.

The major additional computational effort associated with the Newton-Kantorovich method is associated with the evaluation of J'_{mnp} and H'_{mnp} at each iteration. Inspection of (5.51) reveals that, in order to compute $\{J'_{mnp}\}$ and $\{H'_{mnp}\}$ efficiently, it is desirable to use complex exponential functions for $\{\varrho_p(\theta)\}$. A first choice for the representation of $\sigma(\theta)$ might then be

$$\sigma(\theta) = \rho_0 + \sum_{p=1}^{\infty} (\rho_p^{(+)} \exp(jp\theta) + \rho_p^{(-)} \exp(-jp\theta)), \tag{5.56}$$

where the constraints that ρ_0 be real and that $\rho_p^{(+)}$ and $\rho_p^{(-)}$ form complex conjugate pairs must be imposed to ensure that $\sigma(\theta)$ is real-valued. Note that, although J' and H' depend on three indices, the simple addition theorem for complex exponential functions can be exploited to combine the indices p and n into a single index $q = p+n$. Thus, the

complex exponential representation provides for efficient evaluation of J'_{mnp} and H'_{mnp}, because it only involves evaluating integrals of the form

$$H'_{mq} = kC_m \int_0^{2\pi} H_m^{(2)}(k\sigma(\theta)) e^{j(q-m)\theta}\, d\theta, \quad m, q = -\infty, \ldots, \infty. \tag{5.57}$$

The complex representation (5.56) is, however, inconvenient when solving the linear equations from the Newton-Kantorovich method, where the complex conjugacy requirements must be incorporated as constraint equations. Since $\sigma(\theta)$ is a real function of θ, it seems natural to choose a real Fourier series representation

$$\sigma(\theta) = \rho_0 + \sum_{p=1}^{\infty} (\rho_p^{(c)} \cos p\theta + \rho_p^{(s)} \sin p\theta), \tag{5.58}$$

where $\{\rho_p^{(c)}, \rho_p^{(s)}\}$ are real numbers. Fortunately, the advantages of both representations can be preserved by noting the simple transformations between the coefficients :

$$\rho_p^{(c)} = \rho_p^{(+)} + \rho_p^{(-)}, \qquad \rho_p^{(s)} = j(\rho_p^{(+)} - \rho_p^{(-)}), \tag{5.59}$$

$$\rho_p^{(+)} = (\rho_p^{(c)} - j\rho_p^{(s)})/2, \qquad \rho_p^{(-)} = (\rho_p^{(c)} + j\rho_p^{(s)})/2. \tag{5.60}$$

Consequently, the partial derivatives with respect to these quantities transform as

$$\partial/\partial\rho_p^{(c)} = \partial/\partial\rho_p^{(+)} + \partial/\partial\rho_p^{(-)}, \qquad \partial/\partial\rho_p^{(s)} = j(\partial/\partial\rho_p^{(+)} - \partial/\partial\rho_p^{(-)}), \tag{5.61}$$

$$\partial/\partial\rho_p^{(+)} = (\partial/\partial\rho_p^{(c)} - j\partial/\partial\rho_p^{(s)})/2, \qquad \partial/\partial\rho_p^{(-)} = (\partial/\partial\rho_p^{(c)} + j\partial/\partial\rho_p^{(s)})/2. \tag{5.62}$$

Hence a most efficient procedure for computation of Λ' is to first associate $\{\rho_p\}$ with $\{\rho_p^{(+)}, \rho_p^{(-)}\}$, introduced in (5.56), throughout §5.2.3 so that Λ' is computed in terms of the complex exponential basis representation. The coordinate transformations (5.61) are then invoked to yield Λ' with respect to the set $\{\rho_p\} = \{\rho_p^{(c)}, \rho_p^{(s)}\}$, which is introduced in the real Fourier series representation (5.58). Λ' is invoked when the linear equations in the Newton-Kantorovich method are solved.

It is instructive to think of (5.39), (5.41), and (5.42) as summations of coefficients of basis functions. In any computer implementation of the algorithm, finite dimensional spanning sets must be extracted from such bases. While it would be desirable to have analytical means by which to estimate what these dimensions should be to achieve a specific accuracy, the former are invariably chosen by computational experiment, that is by the unimaginative procedure of increasing the dimension of the spanning sets until convergence is manifest.

Utilisation of finite dimensional subspaces ensures that the linear operators, H and J, used for computation of both of the operators Λ and Λ', become finite dimensional matrices. Consequently, all infinite summations are truncated to a finite number of terms, so that H_{mn} and H'_{mnp} have indices m, n, and p varying from, say, $-N_U$ to N_U, $-N_f$ to N_f, and $-N_\rho$ to N_ρ respectively. To determine $\{f_n\}$ in (5.39), it is necessary that $N_U \geq N_f$. For J_{mn} and J'_{mnp}, m need only vary between $-N_b$ and N_b where $N_b < N_U$. A finite number $N_{\mathcal{M}}$ of data is also postulated so that Λ maps from $\mathbf{R}^{2N_\rho+1}$ into $\mathbf{R}^{N_{\mathcal{M}}}$. Clearly, in order to pose a meaningful inverse problem, it is necessary that $N_{\mathcal{M}} > 2N_\rho$.

The linear systems defined by the matrix operator H, namely (5.39) and (5.52), and the Newton-Kantorovich update equation are inverted with the aid of the singular-valued decomposition (SVD) algorithm (Businger and Golub, 1969). Although this

algorithm is more computationally expensive than, say, Gaussian elimination, its capacity to solve over-determined systems and to provide accurate estimation of the vector 2-norm (see §2.2.9) condition number of a matrix (Golub and Van Loan, 1983, Chapter 1) justifies its use. It should also be noted that the effort expended on computing the SVD is only a small part of that required in the complete implementation of the algorithm. This implementation can be claimed to be superior to previous implementations (Roger, 1981; Colton, 1984; Colton and Monk, 1986; Kristensson and Vogel, 1986) (see §4.3.3). The greatest computational effort is that needed to evaluate H, J, H' and J' by numerical quadrature.

Remember that the solution of (5.52) does not require further matrix inversion because H^{-1} is effectively available from the direct problem, in the form of the SVD of H. By employing the computational procedure described in the paragraph following (5.61), evaluation of Λ' only entails evaluation of H' and J', and multiplication by previously computed quantities.

The major component of the additional computational effort required to obtain an update is in the evaluation of J' and H' at each iteration, which is comparable to solving the direct problem. However, these quantities are used repeatedly for different components of Λ'. Thus, unlike a previous method (Kristensson and Vogel, 1986), this scheme calculates Λ' efficiently when the number of terms used to represent the obstacle is increased.

Note also, that computation of Λ' involves quantities that can, in principle, be evaluated to arbitrary accuracy. Numerical inaccuracies are inevitable, but the numerical conditioning of the evaluation of Λ' is governed by the conditioning of H, and so corresponds to that of solving the direct problem. Any noise in the data for the inverse problem has no effect on the computation of Λ', but of course affects solutions of the linear equations in the Newton-Kantorovich algorithm.

Numerical tests on the portion of the program that solves direct problems shows good agreement with the analytical results for cylinders of constant radii, and for other simple geometries described by Bowman *et al.* (1969, §3.2.1). The operator H is truncated to a square matrix of rank N_f. Choosing $N_U = N_f = 15$ and $N_b = 10$, was sufficient to satisfy the forward scattering theorem (see §2.2.10) to a relative accuracy of no worse than 10^{-3} each time a direct problem is solved.

Inverse problem data are generated by computing the normalised far-field scattering data (as defined in §2.2.10) for an incident field e^{-jkx}, invoking either the Dirichlet or Neumann boundary condition, for an obstacle described in dimensionless terms by

$$\frac{2\pi}{\lambda}\sigma(\phi) = 2 + \alpha\cos(4\theta), \tag{5.63}$$

where α is set to 0.5 (see fig. 5.7). For the inverse problem, N_ρ is set to 4 and $N_{\mathcal{M}}$ to 9. The scattering data is specified at angles in the far-field evenly spaced by $2\pi/9$ radians, with $\phi = 0$ being one of the angles. The real and imaginary parts of the scattered field are treated separately, so that the linear equations in the Newton-Kantorovich algorithm become an over-determined system of 18 equations in $2N_\rho + 1 = 9$ unknowns. Such over-determination is essential in practical situations where noise is inevitably present. Since the principal aim here is to demonstrate the validity of the form of the Fréchet derivative introduced in §5.2.3, no noise was added to the data. The Newton-Kantorovich algorithm can be modified to account for noise by replacing $\Lambda(\sigma) = \psi$

Boundary condition	Iteration	Direct problem condition number	Inverse problem condition number	L_∞ error after update
Dirichlet	0			0.5
	1	1.0	25.4	0.130
	2	10.4	38.9	0.025
	3	87.0	40.0	0.000 14
	4	69.9	38.8	0.000 005 6
Neumann	0			0.5
	1	1.0	9.3	0.19
	2	11.5	9.2	0.019
	3	10.0	8.4	0.000 25
	4	9.5	7.5	0.000 000 36

Table 5.6: Results for reconstructions of scatterers introduced in §5.2.4. The Dirichlet boundary condition applies to the scatterer defined by $\frac{2\pi}{\lambda}\sigma(\theta) = 2+0.5\cos(4\theta)$; the Neumann boundary condition applies to the scatterer defined by $\frac{2\pi}{\lambda}\sigma(\theta) = 2 + 0.3\cos(4\theta)$. The data are specified as the complex amplitude of the far-scattered field.

by an equivalent optimisation together with appropriate regularising constraints. One way of incorporating the constraints is by utilizing a penalty function (Kristensson and Vogel, 1986). The Fréchet derivative evaluated as described in §5.2.3 and §5.2.4 can be employed directly in the penalty function method, as described by Kristensson and Vogel (1986), because the standard Gauss-Newton approximation to the Hessian matrix is utilised.

The initial estimate of the obstacle is chosen arbitrarily to be the circle $\frac{2\pi}{\lambda}\sigma(\theta) = 2$, shown as the dashed curve in fig. 5.7. The final reconstruction is visually indistinguishable from the true obstacle. Table 5.6 shows results for 4 iterations of the Newton-Kantorovich method, for both the Dirichlet and Neumann boundary conditions. The conditioning of the direct problem and of the updating of the linear equations in the Newton-Kantorovich method solved at each iteration are listed, as is the error in the estimate of the obstacle expressed as the relative infinity norm (see §2.2.9). The error is seen to exhibit quadratic convergence, demonstrating the effectiveness of the algorithm. To achieve convergence for the Neumann example α has to be reduced to 0.3.

When the scattering data are defined as the bistatic cross-section (see §2.2.10), rather than the complex amplitude, of the normalised scattered field, it is to be expected that the linear approximations invoked at each Newton-Kantorovich iteration can only be valid over a smaller range. Accordingly, bistatic cross-sections have been computed for an object with α set to 0.1 in (5.63). As expected, it transpired that, although $\frac{2\pi}{\lambda}\sigma(\theta) = 2$ was an adequate initial estimate to ensure convergence when the Neumann boundary condition was applied, the Dirichlet boundary condition required an initial estimate $\frac{2\pi}{\lambda}\sigma(\theta) = 2 + 0.05\cos(4\theta)$. The algorithm's performance is summarised in table 5.7.

Further results are summarised in table 5.8, wherein the results listed in table 5.6

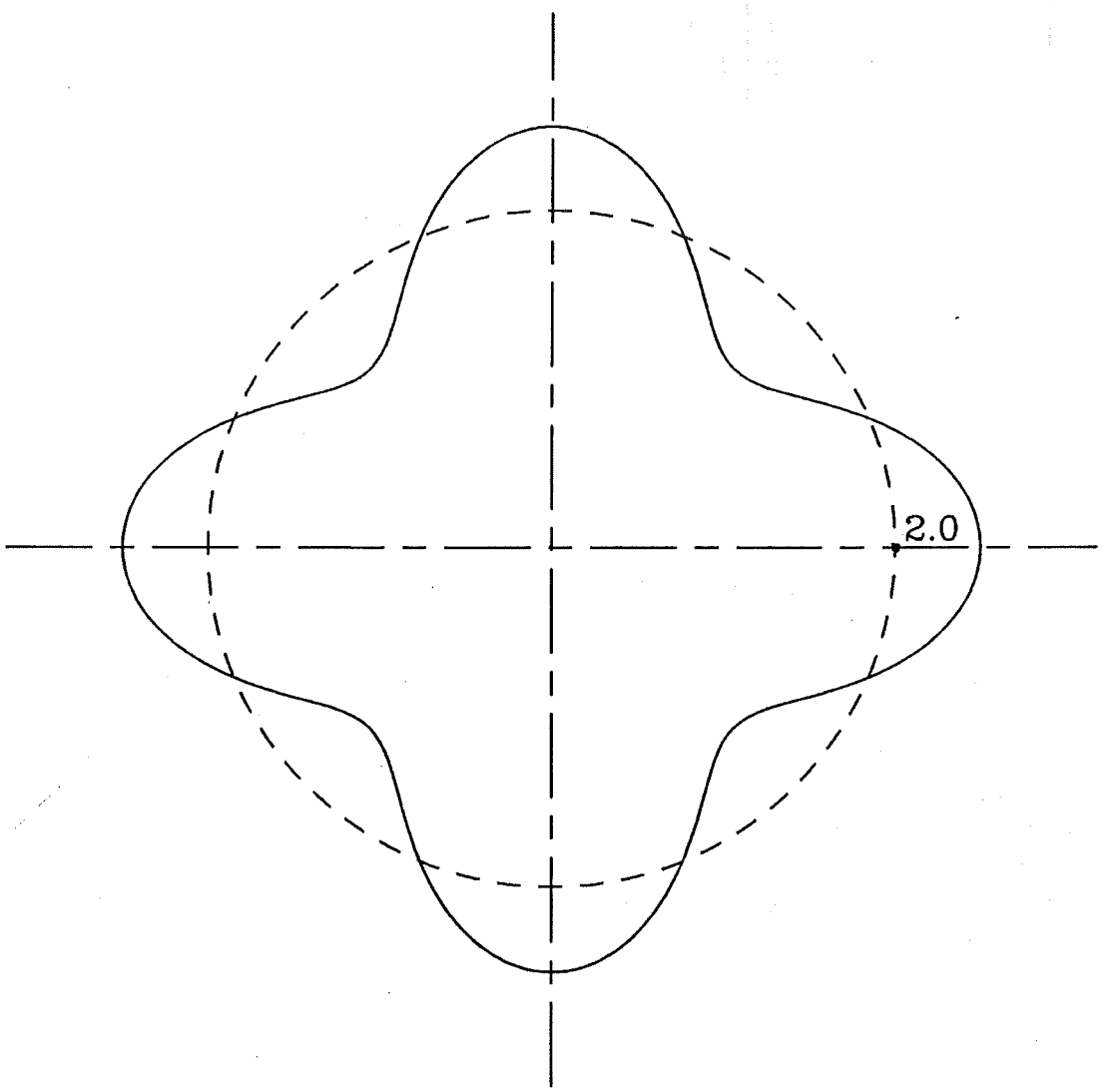


Figure 5.7: Geometry of the scatterer introduced in §5.2.4 defined by $\frac{2\pi}{\lambda}\sigma(\theta) = 2 + 0.5\cos(4\theta)$. The solid line is the boundary of the scatterer, and the dashed circle is the initial estimate of $\sigma(\theta)$.

Boundary condition	Iteration	Direct problem condition number	Inverse problem condition number	L_∞ error after update
Dirichlet	0			0.05
	1	1.5	304.2	0.046
	2	3.3	137.4	0.001 6
	3	2.3	149.8	0.000 043
	4	2.3	150.0	0.000 019
Neumann	0			0.1
	1	1.0	1.0×10^6	0.028
	2	2.2	165.6	0.011
	3	2.4	87.1	0.000 18
	4	2.2	89.3	0.000 000 72

Table 5.7: Results for reconstructions of the scatterer introduced in §5.2.4, having the boundary defined by $\frac{2\pi}{\lambda}\sigma(\theta) = 2 + 0.1\cos(4\theta)$, when the data are specified as the intensity of the far-scattered field.

Object	Boundary Condition	Scattering data type (see §2.2.10)	Error	
			ϵ	Number of iterations
$\sigma(\theta) = \frac{\lambda}{\pi} + \frac{\lambda}{4\pi}\cos 4\theta$	Dirichlet	Complex	2.8×10^{-6}	4
$\sigma(\theta) = \frac{\lambda}{\pi} + \frac{\lambda}{6\pi}\cos 4\theta$	Neumann	Complex	1.8×10^{-7}	4
$\sigma(\theta) = \frac{\lambda}{\pi} + \frac{\lambda}{40\pi}\cos 4\theta$	Dirichlet	Bistatic	9.3×10^{-6}	4
$\sigma(\theta) = \frac{\lambda}{\pi} + \frac{\lambda}{40\pi}\cos 4\theta$	Neumann	Bistatic	3.58×10^{-7}	4
$\sigma(\theta) = \frac{2\lambda}{\pi} + \frac{\lambda}{2\pi}\cos 4\theta$	Dirichlet	Complex	9.0×10^{-7}	5
$\sigma(\theta) = \frac{\lambda}{2\pi} + \frac{\lambda}{8\pi}\cos 4\theta$	Dirichlet	Complex	5.0×10^{-7}	5
$\sigma(\theta) = \frac{\lambda}{\pi} + \frac{\lambda}{4\pi}\cos 4\theta$	Dirichlet	Complex*	5.0×10^{-7}	6
$\sigma(\theta) = \frac{\lambda}{2\pi} + \frac{\lambda}{4\pi}\cos\theta$	Dirichlet	Complex	1.4×10^{-4}	9
$\sigma(\theta) = \frac{\lambda}{\pi} + \frac{\lambda}{4\pi}\cos\theta$	Dirichlet	Complex	1.0×10^{-6}	5
$\sigma(\theta) = \frac{\lambda}{\pi} + \frac{\lambda}{4\pi}\cos\theta$	Neumann	Complex	1.0×10^{-6}	5
$\sigma(\theta) = \frac{\lambda}{2\pi} - \frac{\lambda}{4\pi}\cos 2\theta$	Dirichlet	Complex	6.5×10^{-2}	5
$\sigma(\theta) = \frac{\lambda}{\pi} - \frac{\lambda}{4\pi}\cos 2\theta$	Dirichlet	Complex	5.0×10^{-6}	5

Table 5.8: Summary of the performance of the algorithm introduced in §5.2.2 through §5.2.4. One incident field is invoked for all examples. The scattering data are specified at 7 equally spaced angles in the angular interval $0° - 360°$ except for the entry marked with an astrix. For this entry the scattering data are specified at 7 equally spaced angles in the angular interval $0° - 180°$. The error is specified as a relative infinity norm error (see §2.2.9).

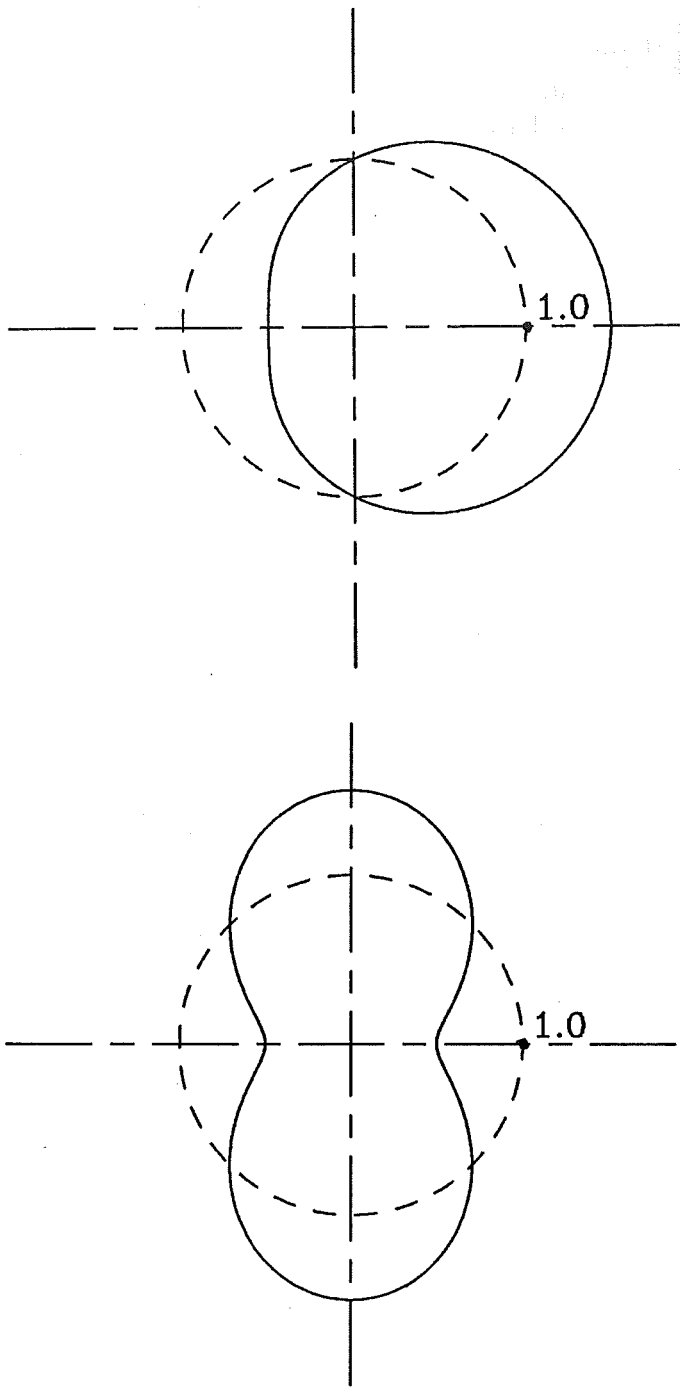


Figure 5.8: Geometry of the scatterers introduced in rows 9 and 11 of table 5.8. The boundaries of the upper and lower scatterers are defined respectively, by $\frac{2\pi}{\lambda}\sigma(\theta) = 1.0 + 0.5\cos(\theta)$ and $\frac{2\pi}{\lambda}\sigma(\theta) = 1.0 - 0.5\cos(2\theta)$. The dashed circles are the initial estimates of $\sigma(\theta)$.

and table 5.7 are also included. These examples are included here to allow comparisons to be made with the methods for reconstructing $\sigma(\mathbf{x})$ presented in §4.3.3. The results indicate that the method introduced in this section (§5.2) compares favourably, since all of the results presented here relate to scattering data being specified at only seven angles and for only a single incident field. However, for the objects, for which the size of Υ_{-+} is small compared to that of Υ_{-}, the stability of the null field solution to the direct problem is reduced. Rows 9 and 11 of table 5.8 highlight this point, which bears on part of what is discussed in §3.3.

5.3 An Extended Rayleigh-Gans (Born) Approximation

This section introduces an original explicit approximate solution to the inverse problem, here called the extended Rayleigh-Gans (Born) approximation. The solution represents an attempt to overcome some of the limitations associated with the Rayleigh-Gans (Born) and Rytov approximations. The solution invokes concepts introduced for the high frequency approximation (see §3.1), the volume source formulation (see §3.2) and also the Rayleigh-Gans (Born) approximation (see §3.2.1).

Alternative explicit approximate solutions have been devised in an attempt to overcome limitations associated with the Rayleigh-Gans (Born) and Rytov approximations. However, these methods either require *a priori* information about the object (Devaney and Oristaglio, 1983; Beylkin, 1985a), or when no *a priori* information is required the few numerical results so far presented do not suggest that the improvements are likely to be spectacular (Dunlop *et al.*, 1976; Bates *et al.*, 1976; Dunlop, 1978; Soumekh, 1986).

In §5.3.1 the extended Rayleigh-Gans (Born) approximation is formally introduced, and in §5.3.2 its implementation is described. In §5.3.3 numerical results are presented. These results are compared to what are obtained when the conventional Rayleigh-Gans (Born) approximation is invoked in order to show that the accuracy of reconstruction has been improved.

5.3.1 Formal Solution

Invoking the volume source formulation (3.28) and the far-field form of the Green's function (see §2.2.4), the scattered field can be expressed as

$$\psi_{ff_s}(\hat{\mathbf{x}}, k) = \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x}_1) - 1)\psi(\mathbf{x}_1, k)e^{jk\hat{\mathbf{x}}\cdot\mathbf{x}_1}\, d\Upsilon(\mathbf{x}_1) \tag{5.64}$$

To make (5.64) more manageable, $\psi(\mathbf{x}, k)$ is represented by the high frequency approximation (3.4) given in §3.1. This results in

$$\psi_{ff_s}(\hat{\mathbf{x}}, k) = \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x}) - 1)e^{-jkS(\mathbf{x}_1)}e^{jk\hat{\mathbf{x}}\cdot\mathbf{x}_1} d\Upsilon(\mathbf{x}_1) \tag{5.65}$$

To take advantage of the numerical efficiency of the FFT algorithm (Cooley and Tukey, 1965; Brigham, 1974), (5.65) is manipulated into the form of a Fourier integral. By appealing to the procedure invoked for the Rayleigh-Gans (Born) approximation (see §3.2.1), (5.65) is transformed to

$$\psi_{ff_s}(\hat{\mathbf{x}}, k) = \int_{\Upsilon_-} k^2(\chi^2(\mathbf{x})-)e^{-jkS(\mathbf{x}_1)}e^{j\mathbf{k}\cdot\mathbf{x}_1}e^{-j\mathbf{k}\cdot\mathbf{x}_1}e^{jk\hat{\mathbf{x}}\cdot\mathbf{x}_1} d\Upsilon(\mathbf{x}_1) \tag{5.66}$$

On defining

$$f(\mathbf{x},\hat{\mathbf{k}}) = [\chi^2(\mathbf{x}) - 1]e^{-jk(S(\mathbf{x})-\hat{\mathbf{k}}\cdot\mathbf{x})} \tag{5.67}$$

(5.66) can be rewritten as

$$\begin{aligned}\psi_{ff_s}(\hat{\mathbf{x}},k) &= \int_{\Upsilon_-} f(\mathbf{x}_1,\hat{\mathbf{k}})e^{-jk(\hat{\mathbf{k}}-\hat{\mathbf{x}})\cdot\mathbf{x}_1}d\Upsilon(\mathbf{x}_1) \\ &= k^2\mathcal{F}_2\left[f(\mathbf{x},\hat{\mathbf{k}})\right](\mathbf{k}')\end{aligned} \tag{5.68}$$

where $\mathbf{k}' = k(\hat{\mathbf{k}} - \hat{\mathbf{x}})$. The relationship between the quantity $\psi_{ff_s}(\hat{\mathbf{x}},k)$ and $f(\mathbf{x},\hat{\mathbf{k}})$ is equivalent to the relationship between $\psi_{ff_s}(\hat{\mathbf{x}},k)$ and $\chi(\mathbf{x})$ in the Rayleigh-Gans (Born) approximation when $\hat{\mathbf{k}}$ is fixed. Recall from §3.2.1 that the locus of points given by $\mathbf{k}'$ for constant $\hat{\mathbf{k}}$ is a circle whose origin is at the extremity of the vector $\mathbf{k}'$. For different $\hat{\mathbf{k}}$, however, the quantity $f(\mathbf{x},\hat{\mathbf{k}})$ changes.

In order to reconstruct $\chi(\mathbf{x})$ from $f(\mathbf{x},\hat{\mathbf{k}})$, the quantity

$$D(\mathbf{x},\hat{\mathbf{k}}) = e^{-jk(S(\mathbf{x})-\hat{\mathbf{k}}\cdot\mathbf{x}_1)} \tag{5.69}$$

has to be estimated. $D(\mathbf{x},\hat{\mathbf{k}})$ is here termed the distorting function and compensates for the difference between the incident field and the high frequency approximation to the field in Υ_-. For situations in which the Rayleigh-Gans (Born) approximation is valid, the distorting function is unity. However, for other situations some technique must be devised for estimating $D(\mathbf{x},\hat{\mathbf{x}})$ so that (5.68) can be incorporated into an inverse method. In order to do this, $k(S(\mathbf{x}) - \hat{\mathbf{k}}\cdot\mathbf{x})$ is represented as a power series:

$$k(S(\mathbf{x}) - \hat{\mathbf{k}}\cdot\mathbf{x}) = k(a_0 + \mathbf{a}_1\cdot\mathbf{x} + (\mathbf{a}_2\cdot\mathbf{x})^2 + \cdots) \tag{5.70}$$

where the constants $\mathbf{a}_n$, for $n = 0, 1, \cdots, \infty$ depend on $\hat{\mathbf{k}}$. (5.68) can then be expressed as

$$\psi_{ff_s}(\mathbf{x},k) = k^2e^{jka_0}\mathcal{F}_2\left[\chi^2(\mathbf{x} - k\mathbf{a}_1) - 1\right] \odot \mathcal{F}_2\left[e^{-jk((\mathbf{a}_2\cdot\mathbf{x})^2+\cdots)}\right](\mathbf{k}') \tag{5.71}$$

Then, once the constants $\mathbf{a}_n$ have been estimated, the distorting function can be determined in the manner described below in §5.3.2.

5.3.2 Estimating the Distorting Function

To estimate the form of $D(\mathbf{x},\hat{\mathbf{k}})$ it is assumed that the constants a_0 and $\mathbf{a}_1$ are most significant, with the other $\mathbf{a}_n$ being small enough to be ignored. This allows (5.71) to be written as

$$\psi_{ff_s}(\hat{\mathbf{x}},k) = k^2e^{jka_0}\mathcal{F}_2\left[\chi^2(\mathbf{x} - k\mathbf{a}_1) - 1\right](\mathbf{k}') \tag{5.72}$$

If the incident field travels in the direction $\hat{\mathbf{k}}$, and if $\psi_{ff_s}(\hat{\mathbf{x}},k)$ is specified for all $\hat{\mathbf{x}}$, then the extremity of the vector $\mathbf{k}' = k(\hat{\mathbf{k}} - \hat{\mathbf{x}})$ traces out a circle, as illustrated in fig. 5.9. The origin of the circle is at the extremity of the vector $\mathbf{k} + k\mathbf{a}_1$. Also, the spectrum of $(\chi^2(\mathbf{x}) - 1)$ is multiplied by e^{jka_0}. The locus of points relevant to the Rayleigh-Gans (Born) is a circle traced out by the extremity of the vector $\mathbf{k}' = k(\hat{\mathbf{k}}-\hat{\mathbf{x}})$. The origin of the circle is $\mathbf{k}$. This locus is also shown in fig. 5.9 for comparison.

The technique adopted here for estimating the distorting function is based upon calculating $\psi_{ff_s}(\hat{\mathbf{x}},k)$ for two different magnitudes of $\mathbf{k}$, k_1 and k_2 say. For $n = 1$ and $n = 2$, the origin of the circle is translated by $k_n\mathbf{a}_1$ and the radius of the circle

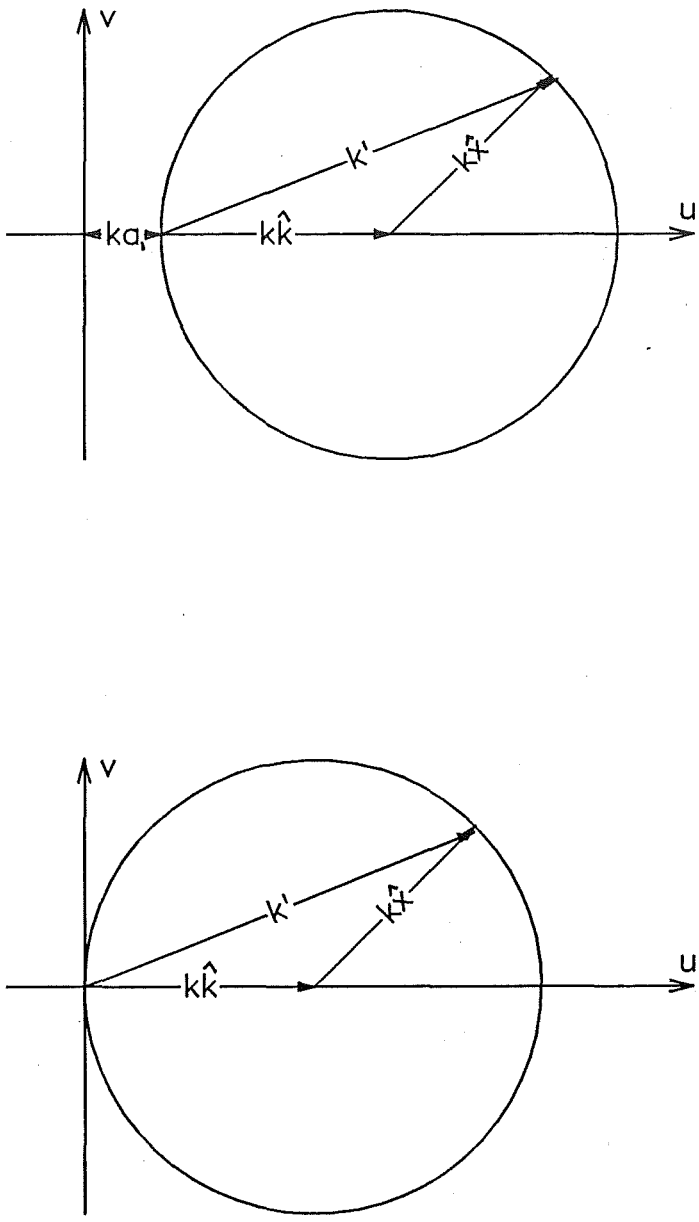


Figure 5.9: Locus in Fourier space corresponding to observable scattered field. The upper and lower loci correspond respectively to the extended Rayleigh-Gans (Born) and conventional Rayleigh-Gans (Born) approximations. The scattered field is related to the parts of the Fourier transform of $k^2(\chi^2(\mathbf{x}) - 1)$ on a circle described by the extremity of the vector $\mathbf{k}' = k(\hat{\mathbf{k}} - \hat{\mathbf{x}})$ for constant $\hat{\mathbf{k}}$ and for all $\hat{\mathbf{x}}$ in Fourier space. The origins of the circles for the extended and conventional Rayleigh-Gans (Born) approximations are, respectively, $\mathbf{k} + k\mathbf{a}_1$ and $\mathbf{k}$.

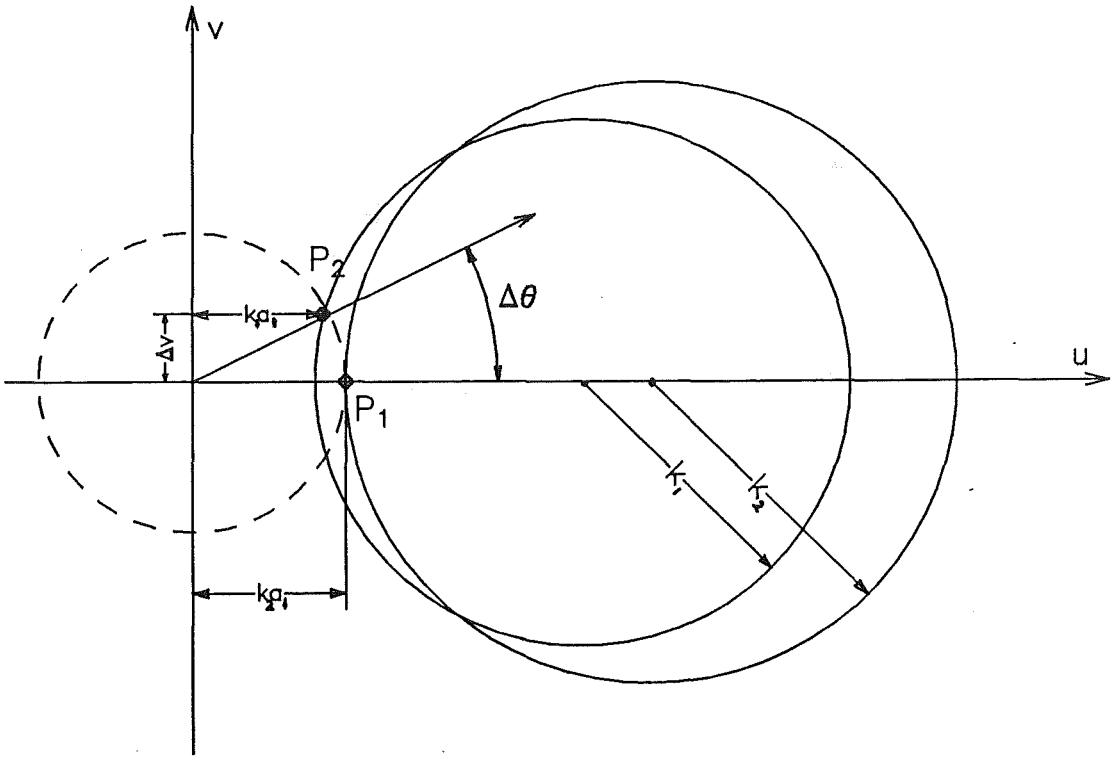


Figure 5.10: Illustration of the procedure invoked in §5.3.2 to estimate $\mathbf{a}_1$. The spectrum of $(\chi^2(\mathbf{x}) - 1)$ is assumed symmetric over intervals $\Delta\theta$, as in defined in (5.73), so that the magnitude of this spectrum has the same value at the points P_1 and P_2. The quantity Δv is identified on the figure because this is the most convenient way to define it.

becomes k_n. Furthermore, the phase $e^{jk_n a_0}$ is added to the spectrum of $(\chi^2(\mathbf{x}) - 1)$. If the maximum extent of Υ_- is L in the spatial domain, then significant variations in the function $\mathcal{F}_2[\chi^2(\mathbf{x}) - 1]$ only occur over $2\pi/L$ in the Fourier domain. Thus, if the values of $\mathbf{k}$ for $n = 1$ and $n = 2$ are such that the loci of points given by $\mathbf{k}'$ for each $\mathbf{k}_n$ are not shifted by more than $2\pi/L$, a_0 can be straightforwardly estimated by obtaining the difference between the phases of $\psi_{ff_s}(\hat{\mathbf{x}}, k_1)$ and $\psi_{ff_s}(\hat{\mathbf{x}}, k_2)$ at any $\hat{\mathbf{x}}$. The difference between k_1 and k_2 can always be made small enough to ensure that this condition holds, since L is always finite.

The constant $\mathbf{a}_1$, however, cannot be estimated without invoking some further approximation. The approximation chosen here is that the spectrum of $(\chi^2(\mathbf{x}) - 1)$ is circularly symmetric over some angular range $\Delta\theta$ so that

$$|F(\alpha;\theta)| \approx |F(\alpha;\theta + \gamma\Delta\theta)| \tag{5.73}$$

for all θ and α, where $\gamma \in [0,1]$ and $F(\alpha;\theta)$ is the Fourier transform of $(\chi^2(\mathbf{x}) - 1)$. If $p(\rho;\theta)$ is the projection at angle θ of $(\chi^2(\mathbf{x}) - 1)$ then by the projection theorem (see §4.2.1.1), (5.73) implies

$$p(\rho;\theta) \approx p(\rho;\theta + \gamma\Delta\theta) \tag{5.74}$$

for all θ and α. It follows that the variation in the projections of $(\chi^2(\mathbf{x}) - 1)$ over $\Delta\theta$ must also be small, provided $\Delta\theta$ is small enough.

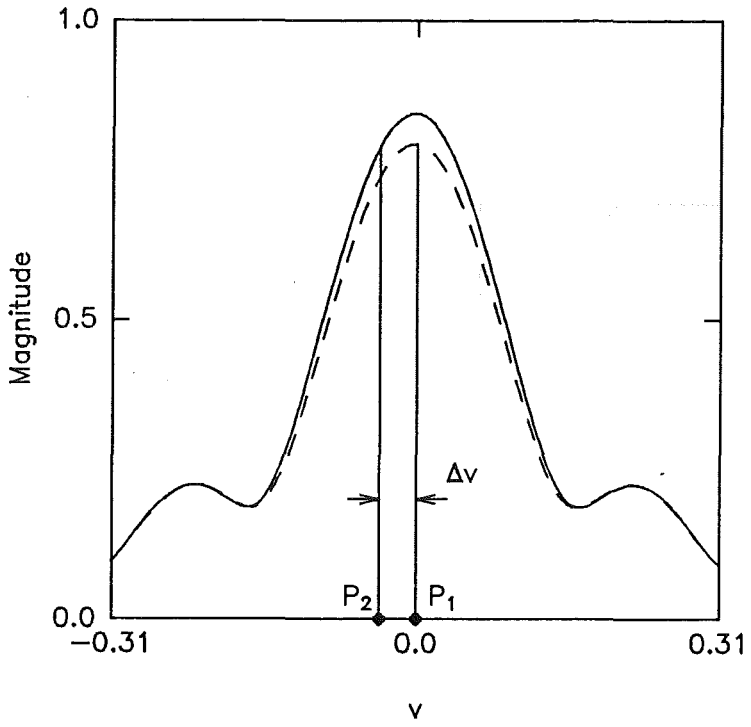


Figure 5.11: Plots of $|\psi_{ffs}(\mathbf{x}, k)|$ versus v, for $k = k_1$ and k_2. The points on the v-axis corresponding to the points P_1 and P_2 shown in fig. 5.10 are identified in this figure.

The method of estimating $\mathbf{a}_1$ is conveniently explained with the aid of fig. 5.10. Since the spectrum of $(\chi^2(\mathbf{x}) - 1)$ is assumed symmetric over angular intervals of width $\Delta\theta$, the magnitude of $\chi^2(\mathbf{x})$ at P_1 is the same as at P_2. This allows the quantity Δv (defined in fig. 5.10) to be calculated. Consequently an estimate of the distance $(k_1 - k_2)\mathbf{a}_1$ can be obtained by algebraic manipulation and hence $\mathbf{a}_1$ can be determined. The angular interval $\Delta\theta$ can be made as small as one desires by invoking k_1 and k_2 arbitrarily close together.

The points P_1 and P_2 are located by plotting $|\psi_{ffs}(\hat{\mathbf{x}}, k)|$ against v as illustrated in fig. 5.11. The maximum of $|\psi_{ffs}(\hat{\mathbf{x}}, k_2)|$ is equated to $|\psi_{ffs}(\hat{\mathbf{x}}, k_1)|$ to find Δv. Once Δv is evaluated, algebraic manipulation allows the vector $(k_1 - k_2)\mathbf{a}_1$ to be determined.

Once a_0 and $\mathbf{a}_1$ are determined, reconstruction can proceed straightforwardly. For each $\hat{\mathbf{k}}$, the locus of $\mathbf{k}'$ is translated by $k\mathbf{a}_1$ and the spectrum is multiplied by e^{jka_0}. The locus is then interpolated onto a rectangular grid by some suitable means such as back-propagation, as described in §4.2.2.3.

5.3.3 Numerical Examples

To illustrate the performance of the algorithm, several numerical examples are presented here. The examples are of reconstructions of the standard object defined in fig. 3.3 for various ν_- and r_0. Reconstructions obtained by invoking the conventional Rayleigh-Gans (Born) approximation are also presented. This allows the significance of the extension to the Rayleigh-Gans (Born) approximation to be assessed.

The first example is for $2r_0 = 6\lambda$ and $\nu_- = 1.05$. The magnitude and phase differences between, $\psi_{ffs}(\hat{\mathbf{x}}, k)$ for $k = 1$ and $k = 1.05$ are plotted as functions of v in

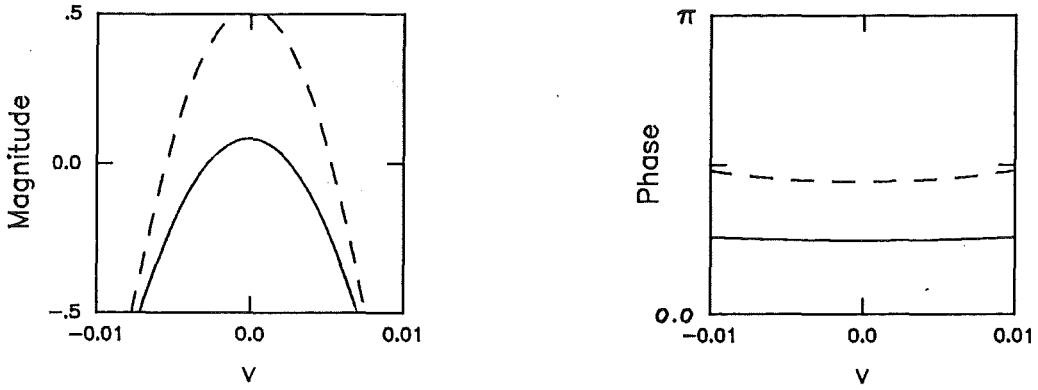


Figure 5.12: Magnitude and phase differences between, $\psi_{ff_s}(\hat{\mathbf{x}}, 1.0)$ and $\psi_{ff_s}(\hat{\mathbf{x}}, 1.05)$ for the standard object plotted against v. The solid lines are for $\nu_- = 1.05$, $2r_0 = 6\lambda$ and the dashed lines are for $\nu_- = 1.10$, $2r_0 = 6\lambda$

fig. 5.12. The phase difference curve indicates that $a_0 = .88$. The zero crossing of the magnitude difference curve indicates that $\mathbf{a}_1 = 0.05$. This actually accords with simple-minded intuition because the extra phase delay imposed on the incident field on its passage to the origin is approximately $(\nu_- - 1)r_0 = 0.94$ and $\mathbf{k} \cdot \mathbf{a}_1$ is simply the perturbation in the background refractive index $k(\nu_- - 1)$ which is 0.05. Because the object is symmetric, a_0 and $\mathbf{a}_1$ are the same for all $\hat{\mathbf{k}}$. The resulting reconstruction is shown in fig. 5.13a. It is clear that the extension to the Rayleigh-Gans (Born) approximation provides improved reconstructions. Further reconstructions of the standard object for various ν_- and r_0 are presented in figs. 5.13b through 5.13d. The upper and lower rows of graphs on each of these figures respectively represent reconstructions generated by the extended and conventional Rayleigh-Gans (Born) approximations. All the results confirm that improvement over the conventional Rayleigh-Gans (Born) approximation has been obtained.

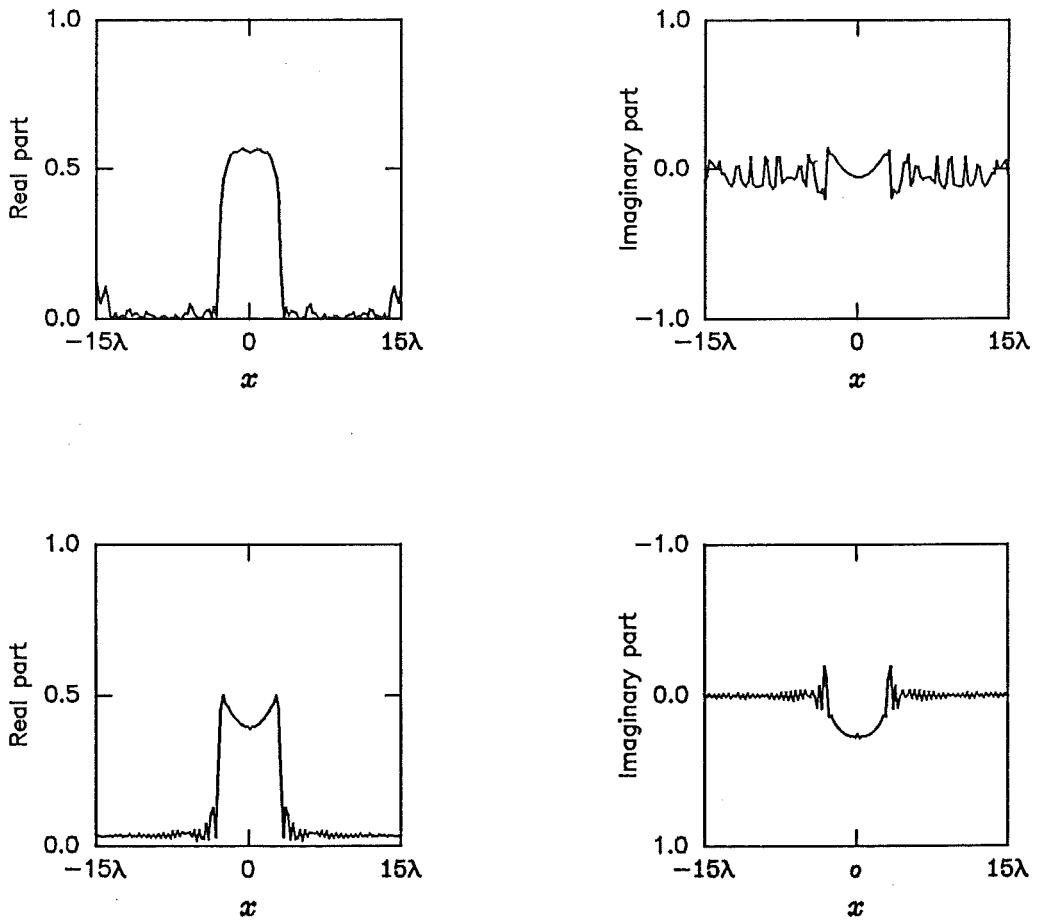


Figure 5.13: Reconstructions of the standard object defined in fig. 3.3. The upper and lower rows of graphs of each sub-figure on each page respectively represent reconstructions generated by the extended and conventional Rayleigh-Gans (Born) approximations: (a) $\nu_- = 1.05$, $2r_0 = 6\lambda$.

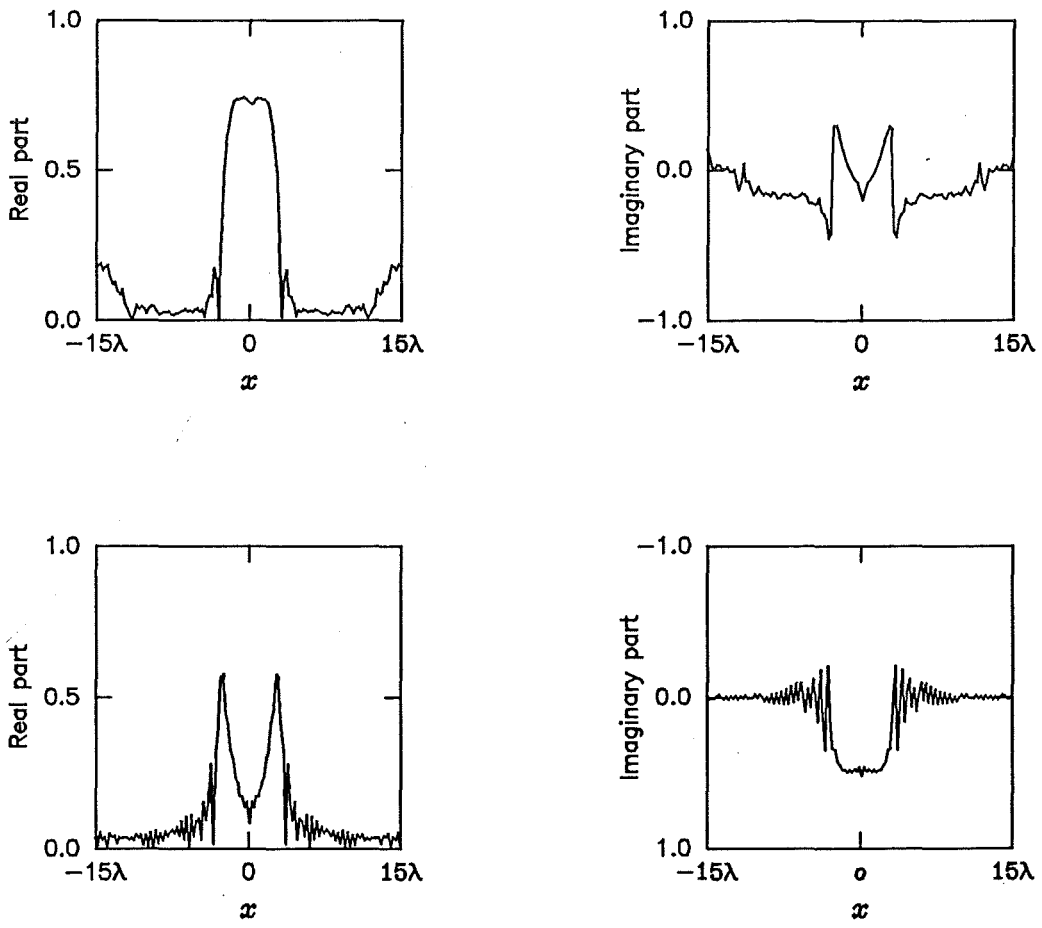


Figure 5.13: (continued) (b) $\nu_- = 1.10$, $2r_0 = 6\lambda$.

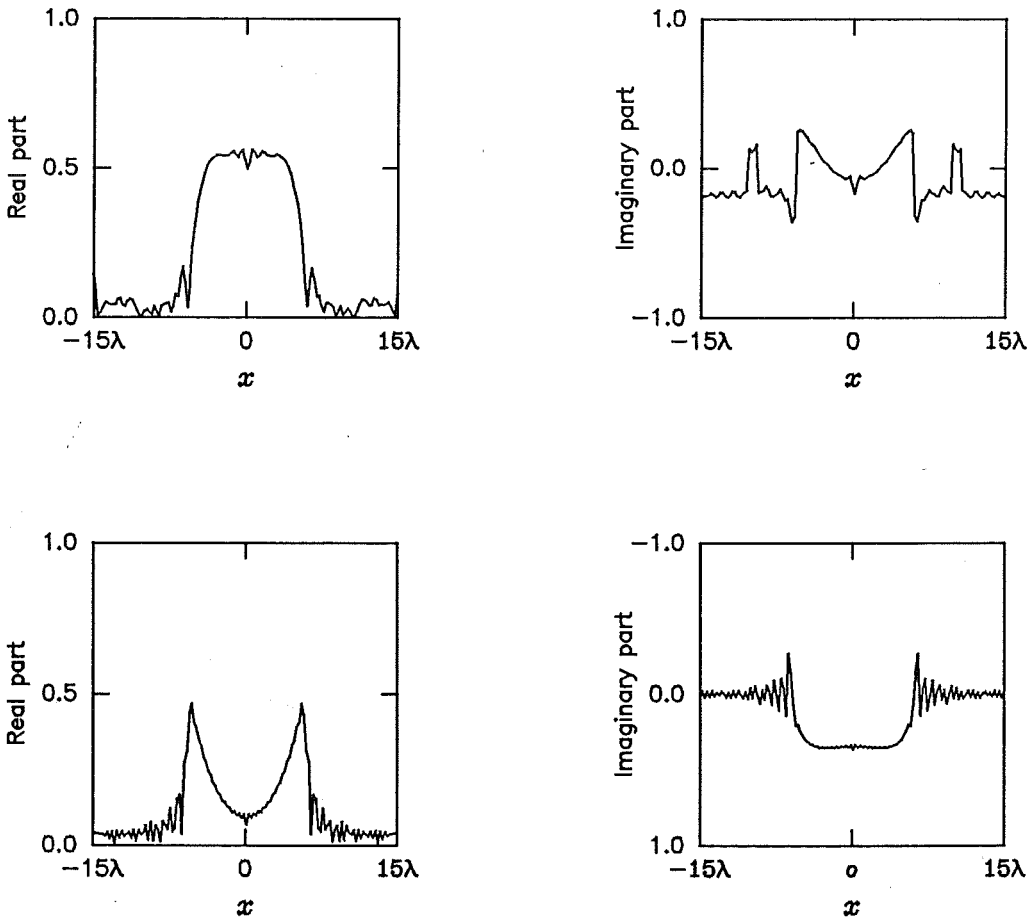


Figure 5.13: (continued) (c) $\nu_- = 1.05$, $2r_0 = 12\lambda$.

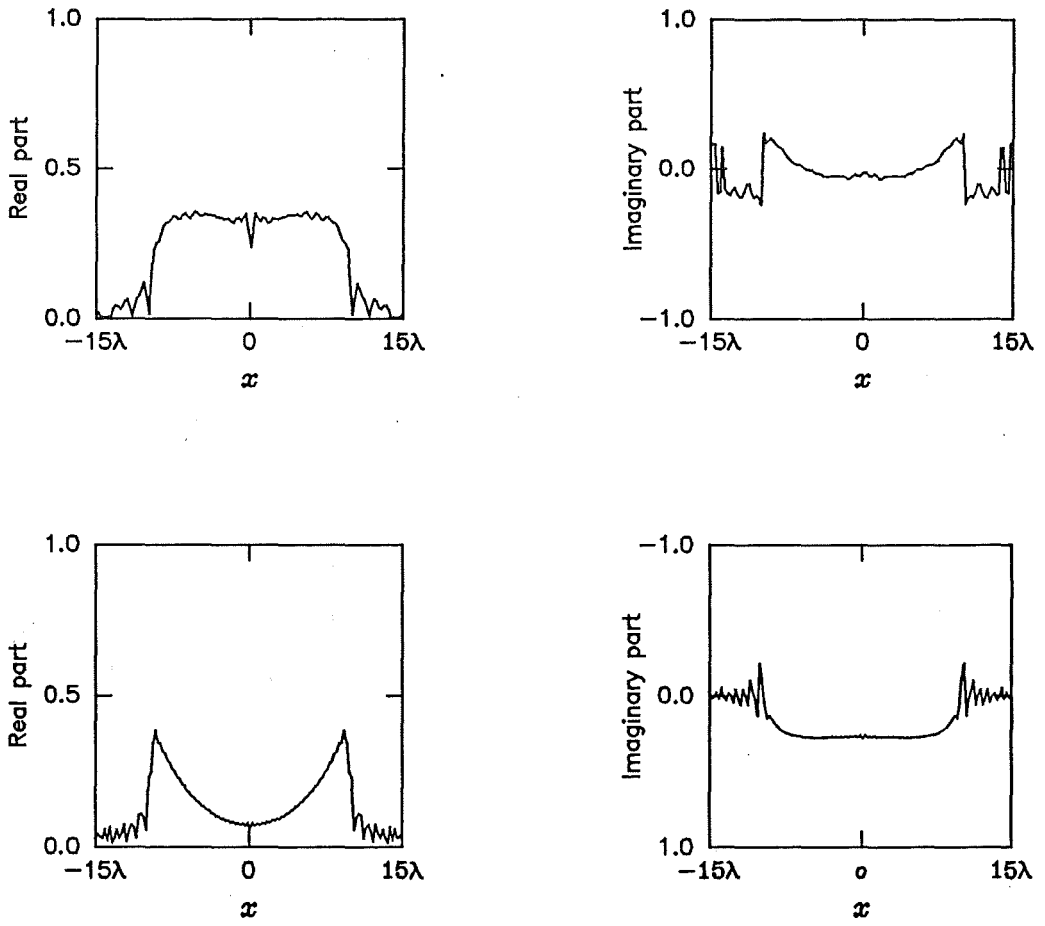


Figure 5.13: (continued) (d) $\nu_- = 1.03$, $2r_0 = 20\lambda$.

Chapter 6

Reconstructing Binary Images from Few Projections

The concern in this Chapter, which expands a recently prepared account (Murch and Bates, 1990a) of an original approach stemming from a series of papers on image reconstruction from projections, the earliest being by Lewitt and Bates (1978) and most recent by Tan *et al.* (1986), is with reconstructing various types of binary images (i.e. those whose pixel values are either zero or unity) from small numbers of projections. This is interesting in its own right because many objects of scientific and technological interest are approximately homogeneous, and even those objects which are palpably inhomogeneous can sometimes be usefully treated as though they are of uniform density. The main concern with binary objects stems, however, from the problem of determining the three-dimensional shape of an object from its silhouettes, the relative orientations of which are unknown *a priori*. While the algorithms introduced in this Chapter are only designed to reconstruct the contours of single cross-sections of objects, they are needed to complement the original approach introduced in Chapter 7 for estimating the shapes of three-dimensional objects from their silhouettes.

After introducing necessary preliminaries in §6.1 and formally stating the problem this Chapter is concerned with, two theorems central to solving this problem are presented in §6.2. This is followed in §6.3 by the algorithm for reconstructing a convex cross-section from a pair of projections, supplemented in §6.4 by illustration of degradations due to the given projections being unavoidably imperfect in practice and the ameliorative effect of employing more than two projections. By appealing to an established technique (Garden and Bates, 1984) it is indicated in §6.5 how to initially estimate the contour of a non-convex, or multiple-connected, cross-section. §6.6 describes a modified algorithm for reconstructing non-convex cross-sections from small numbers of projections.

Illustrative examples are invoked in §6.3 through §6.6 to demonstrate the efficacy of the algorithms. The significance of the results reported in this chapter are assessed in Chapter 8, which also indicates the further work that needs to be done to increase the usefulness of the approach.

6.1 Preliminaries

The coordinate systems shown in fig. 6.1 are employed. The closed curve $\sigma(\mathbf{x})$ is the perimeter of an arbitrary cross-section of an arbitrary body, with Υ_- and Υ_+ being the regions of the x, y-plane respectively inside and outside $\sigma(\mathbf{x})$. Since bodies composed of distinct, separated parts are studied in §6.5 and §6.6, it is envisaged that $\sigma(\mathbf{x})$ may be replaced by several non-intersecting closed curves, with Υ_- being the union of the interiors of all these curves. The concern herein is with bodies of constant density, which are referred to as binary objects. Normalising the density throughout Υ_- to unity, it is seen that the density $\chi(x, y)$ throughout the x, y-plane is given by

$$\chi(x,y) = \begin{cases} 1 & \text{for} \quad P \in \Upsilon_- \\ 0 & \text{for} \quad P \in \Upsilon_+ \end{cases} \tag{6.1}$$

Reference to (4.9) then confirms that the projection at angle ϕ through the cross-section depicted in fig. 6.1 is

$$p(\xi;\phi) = \int_{L_-(\xi;\phi)} d\eta \tag{6.2}$$

where $L_-(\xi;\phi)$ is the part of the line $L(\xi;\phi)$ lying inside $\sigma(\mathbf{x})$, or inside all the aforesaid bounding curves. It follows from (6.1) and (6.2) that

$$\int_{-\infty}^{\infty} p(\xi;\phi)\, d\xi = A \tag{6.3}$$

for any ϕ, where A is the area of Υ_-. Note that the infinite limits in (6.3) merely emphasise that the integral is over the whole projection.

In this Chapter $p(\xi;\phi)$ is referred to as a perfect projection. Since there is no escaping imperfections in the real-world, it is convenient to define the contaminated projection $p_c(\xi;\phi)$, corresponding to $p(\xi;\phi)$, by

$$p_c(\xi;\phi) = p(\xi;\phi) + c(\xi;\phi) \tag{6.4}$$

where the contamination $c(\xi;\phi)$ incorporates any inadequacies of the simple model represented by (6.2), such as departures from uniformity of the density of the object, as well as measurement noise and the effects of non-linearities of the recording apparatus.

The general problem considered in this chapter can be posed in the following way. Given a set $\{p(\xi;\phi_n) : n = 1, 2, \cdots, N\}$ of projections of a binary cross-section, at angles belonging to the set $\{\phi_n : n = 1, 2, \cdots, N; 0 \leq \phi_1; \phi_N \leq 2\pi; \phi_n < \phi_{n+1} \quad \text{for all} \quad n\}$, reconstruct the shape of the cross-section. The solution to this problem is of course well known when N is large as is customary for computed (or computer-assisted) tomography (Lewitt and Bates, 1978). What is of particular concern here is how to reduce N as much as possible under various circumstances. It is emphasised that the reconstructions are necessarily computed at points on a square grid. The necessarily finite spacing of the grid points defines the size of each pixel in the x, y-plane.

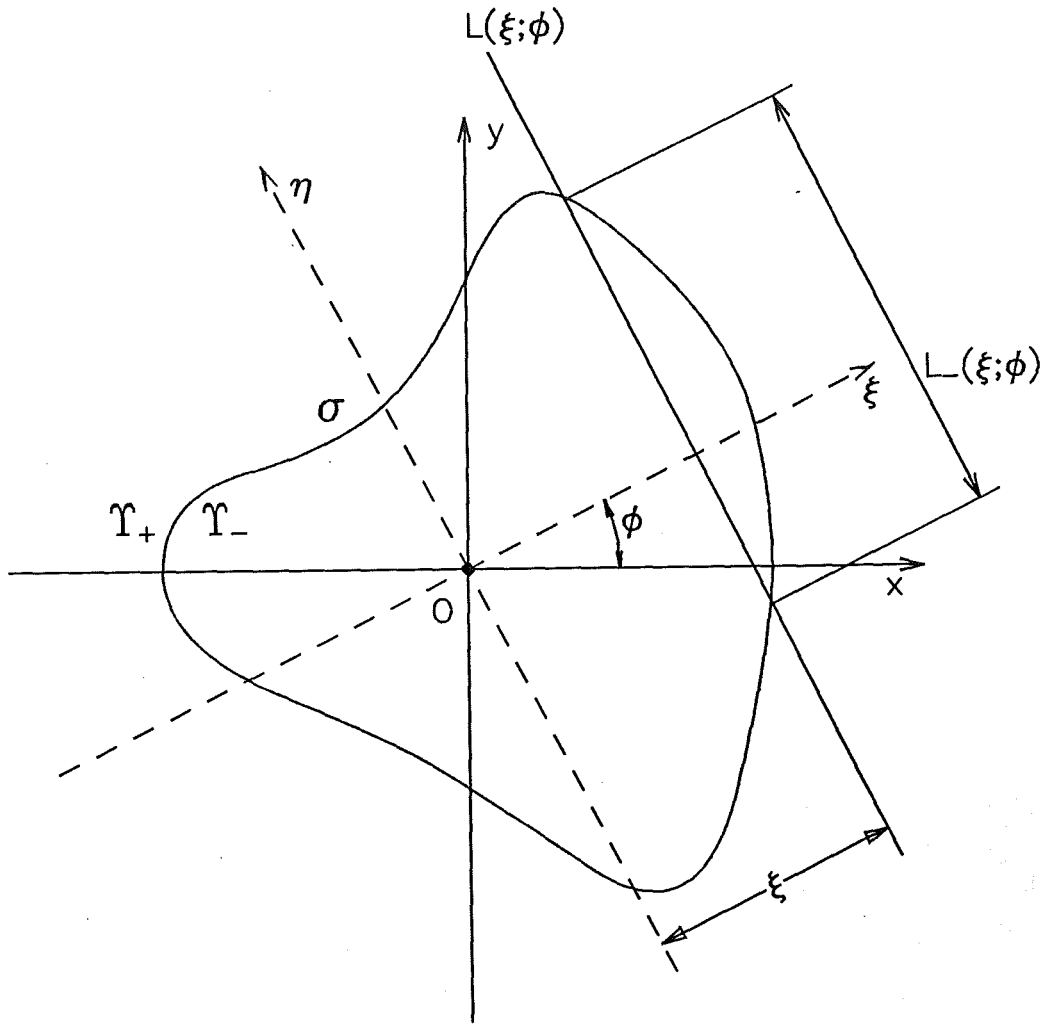


Figure 6.1: Coordinates for cross-section, bounded by the closed curve $\sigma(\mathbf{x})$, of binary object. Cartesian coordinates (x, y) are fixed in cross-section whereas Cartesian coordinates (ξ, η) are rotated by the arbitrary angle ϕ. $L(\xi;\pi)$ is the straight line parallel to, and distant ξ from, the η-axis, with $L_{-}(\xi;\pi)$ being the part of $L(\xi;\pi)$ lying inside $\sigma(\mathbf{x})$. Υ_- and Υ_+ are, respectively, the regions inside and outside $\sigma(\mathbf{x})$.

6.2 Centre-of-Mass and Width Theorems and Registration of Projections

If the origin O of coordinates in fig. 6.1 is at the centre of mass of the cross-section then by definition (Franklin, 1960),

$$\int\int_{(\xi,\eta)\in\Upsilon_-} \chi(x,y)\xi \, d\xi d\eta = 0 \tag{6.5}$$

for any ϕ. It follows from the definition (6.1) and (6.2) that

$$\int_{-\infty}^{\infty} p(\xi;\phi)\xi \, d\xi = 0 \tag{6.6}$$

The two formulas (6.5) and (6.6) constitute what is termed here the centre-of-mass theorem, the crucial aspect of which is that the centre of mass of the projection at any angle ϕ lies on the η-axis, as defined for that ϕ.

It follows from (6.2), for all ξ and ϕ, that

$$p(\xi;\phi) = L_-(\xi;\phi) \tag{6.7}$$

which constitutes what is termed here the width theorem because it confirms that the shape of the cross-section is implicit in a single projection. Reference to fig. 6.1 indicates that $L_-(\xi;\phi)$ is the width of the cross-section along the line parallel to, and distant ξ from, the η-axis. From (6.7), therefore, $p(\xi;\phi)$ also equals this width. The problem is thus seen to reduce to discovering how to rearrange given projections so as to generate the shape of the cross-section.

Fig. 6.2 illustrates how the centre-of-mass theorem permits one to register a pair of projections correctly with respect to each other and to the origin O of coordinates. The ξ-axes of the two projections are oriented with respect to each other at the difference between the angles at which they are defined. Having identified on each ξ-axis the point corresponding to the centre of mass of its respective projection, a straight line perpendicular to the respective ξ-axis is constructed through this point. This centres the projection because the straight line coincides with the η-axis for that projection, implying that the intersection of the straight lines centering the projections necessarily coincides with the origin O of coordinates. Straight lines, parallel to their respective η-axis, drawn through the extremities of the two projections, define a parallelogram enclosing Υ_-, as indicated in fig. 6.2.

Any number of projections can be registered in this way. It is enough to consider a third projection. Having appropriately oriented its ξ-axis with respect to either of the two previous ξ-axes, the third projection is centred and it is shifted parallel to its ξ-axis until its η-axis (i.e. the straight line constructed during the centering process) intersects O.

6.3 Reconstructing a Convex Cross-Section from Two Projections

It is shown here that the perimeter $\sigma(\mathbf{x})$ of a convex cross-section can be reconstructed from a pair of projections. The first step in the reconstruction procedure is to register

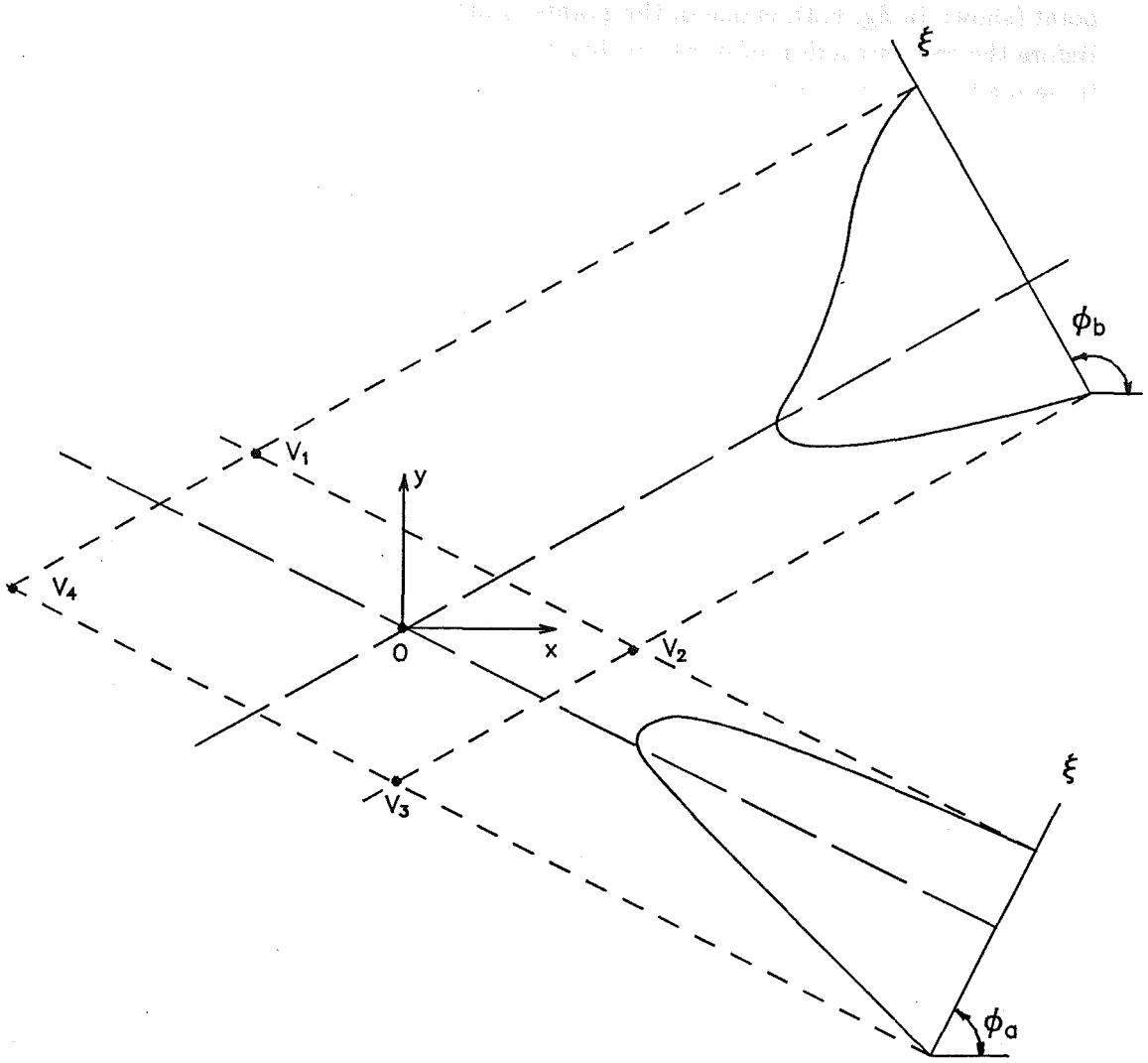


Figure 6.2: Illustration of how two projections, at angles ϕ_a and ϕ_b, are registered with respect to each other by invoking centre-of-mass theorem. Coordinates necessarily coincides with the intersection of the dashed lines, which centre their respective projections. Each pair of dotted lines in the figures is parallel to its respective dashed line. The parallelogram $V_1V_2V_3V_4$ necessarily encloses Υ_-.

the projections (see §6.2 and fig. 6.2). The next step is to estimate the position of a point (called the starting point) on $\sigma(\mathbf{x})$. The definition (6.2) of a projection ensures that $\sigma(\mathbf{x})$ is confined within the parallelogram $V_1V_2V_3V_4$ shown in fig. 6.2. Note that $\sigma(\mathbf{x})$ must touch each side of the parallelogram. Furthermore, if it is convex, $\sigma(\mathbf{x})$ must intersect each of the diagonal lines V_1V_3 and V_2V_4. The point Q_1, chosen as a starting point (shown in fig. 6.3), is one of the points of intersection of $\sigma(\mathbf{x})$ with the line V_1V_3. Before the reconstruction of $\sigma(\mathbf{x})$ can begin, an initial guess at the location of Q_1 has to be made. It is both convenient and effective to take this initial Q_1 to be V_1.

The width theorem (§6.2) ensures that, if Q_1 lies on $\sigma(\mathbf{x})$, then so must the points $Q_{a,1}$ and $Q_{b,1}$ which are shown in fig. 6.3. In keeping with the name of this theorem, each $\overset{(q)}{\longleftrightarrow}$, as defined by illustration in fig. 6.3, is termed a projection width. The width theorem also ensures that the points $Q_{b,a,1}$ and $Q_{a,b,1}$ are located on $\sigma(\mathbf{x})$ with respect to $Q_{a,1}$ and $Q_{b,1}$, respectively as illustrated in fig. 6.3. Similarly, the point $Q_{b,b,2}$ must be located on $\sigma(\mathbf{x})$, as shown, with respect to $Q_{a,b,1}$. Fig. 6.3 displays sufficient additional reconstructed perimeter points Q_q, where q is defined in the caption to fig. 6.3, to unambigously define the scheme for labelling these points. Each time a projection width is invoked to reconstruct another Q_q, it is said that the projection width is taken from the corresponding projection. This procedure is continued until the required number (denoted by K) of perimeter points are reconstructed.

Immediately a Q_q falls outside the parallelogram $V_1V_2V_3V_4$, one knows that the reconstruction has failed due to the location of Q_1 being incorrect. Q_1 is then moved along V_1V_3 one pixel closer to V_3 and the procedure described above is repeated. Q_1 is repeatedly relocated until all Q_q lie inside $V_1V_2V_3V_4$. The extension to N projections of this algorithm is straightforward, for any $N > 2$. Given any Q_q, a further $(N-1)$ perimeter points can be immediately reconstructed from it in the manner illustrated in fig. 6.3. (e.g. as $Q_{a,b,1}$ is generated from $Q_{b,1}$ via the projection width $\overset{(a,b,1)}{\longleftrightarrow}$). Consequently, if J projection widths are taken from each projection, the total number K of perimeter points (excluding Q_1) which are reconstructed is

$$K = 1 + N\sum_{i=1}^{J}(N-1)^{i-1} = 1 + \left[(N-1)^J - 1\right]/(N-2) \tag{6.8}$$

Fig. 6.4a-d shows four different cross-sections, and 65 perimeter points on each of them, reconstructed by the algorithm described above. It is emphasized that the finite size (see final sentence of §6.1) of the pixels prevents the perimeter points from coinciding precisely with $\sigma(\mathbf{x})$.

This algorithm is so efficient computationally that it can tolerate many iterations towards locating the true position (to within whatever accuracy is desired) of Q_1.

6.4 Use of More than Two Projections to Combat Contamination

Once the location of the point Q_1 (defined in fig. 6.3) is correctly identified, all points Q_q (where q is defined in the caption to fig. 6.3) must lie on $\sigma(\mathbf{x})$ to within a tolerance set by the pixel size, provided the given projections are perfect. When the projections are contaminated, however, successive Q_q must wander increasingly from $\sigma(\mathbf{x})$. It is

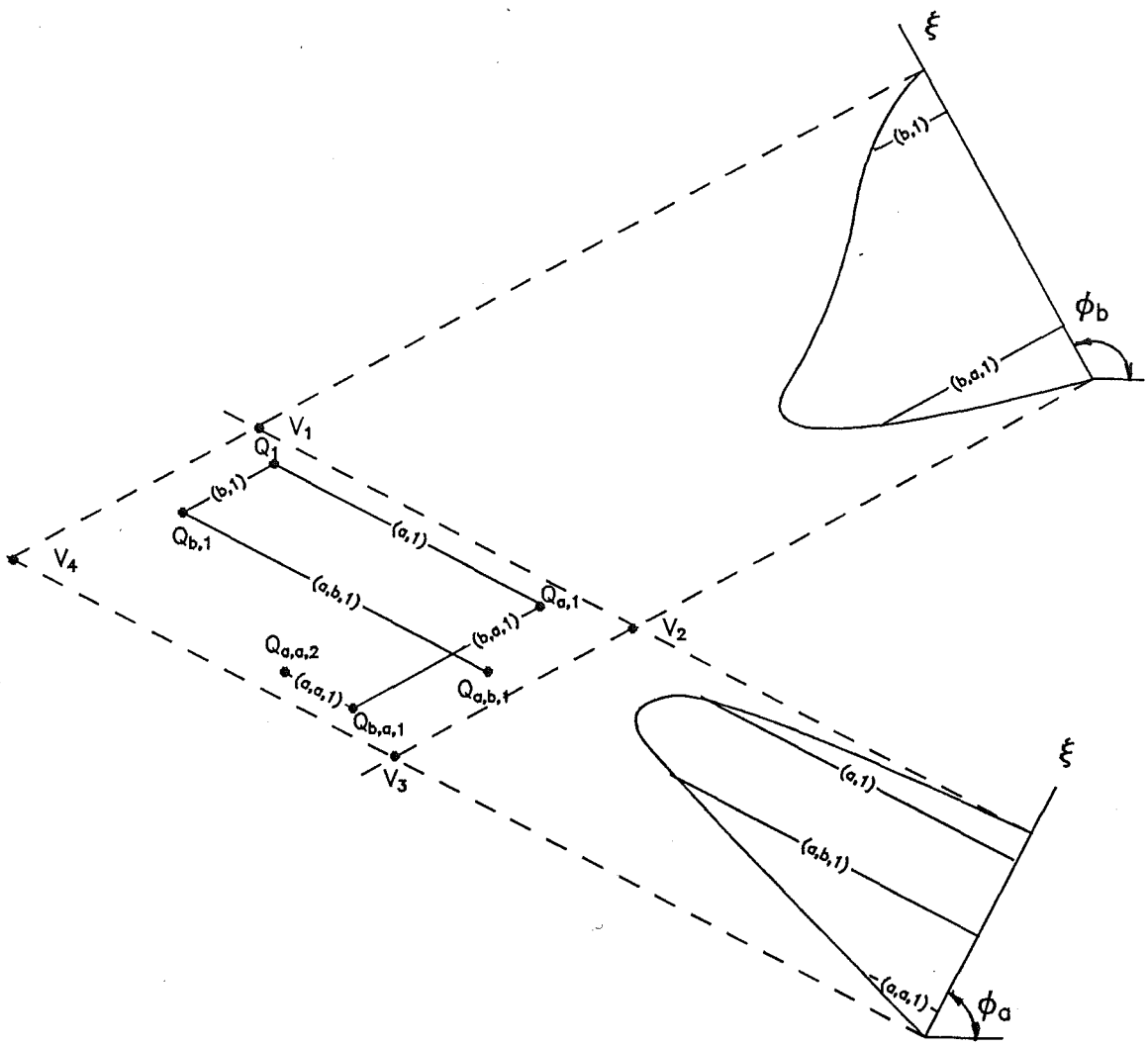
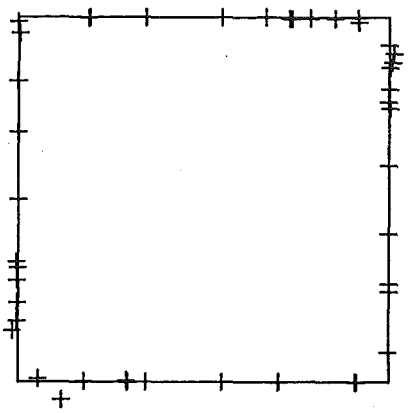
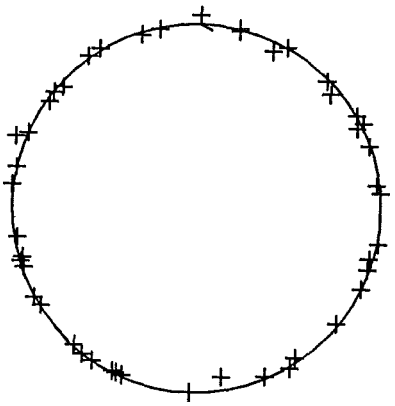


Figure 6.3: Illustration of algorithm for reconstructing convex cross-section from two registered projections (shown in the figure) at angles ϕ_a and ϕ_b. Each pair of $\overset{(q)}{\longleftrightarrow}$ are colinear, where q represents either c, k or c_1, c_2, k, with k being a positive integer, c being either a or b, and c_1, c_2 being either a, b or b, a.

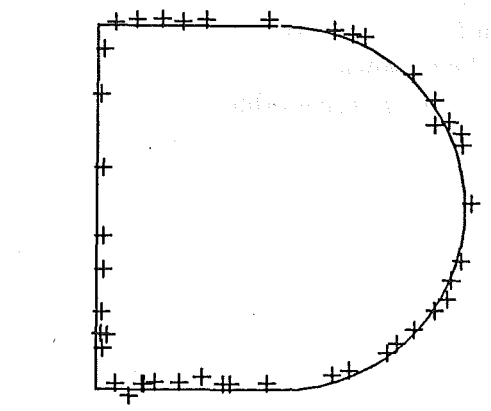


(a)

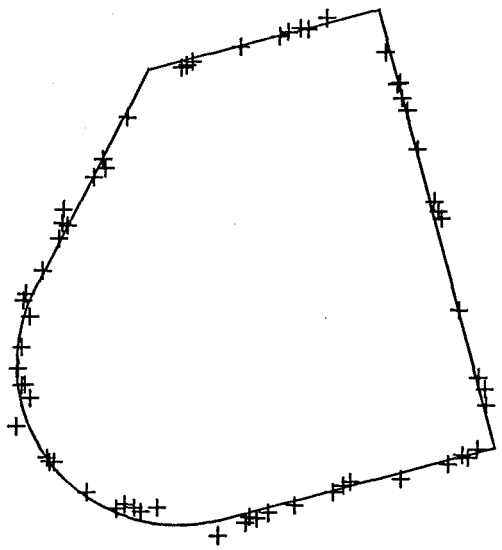


(b)

Figure 6.4: Caption on next page.



(c)



(d)

Figure 6.4: Four convex cross-sections, each reconstructed from a pair of perfect projections, at angles ϕ_a and ϕ_b. Full line represent $\sigma(\mathbf{x})$ and crosses represent reconstructed perimeter points Q_q, which cannot lie exactly on $\sigma(\mathbf{x})$ because of the finite pixel size. The projection angles are: (a) $\phi_a = 45°$, $\phi_b = 130°$ (b) $\phi_a = 68°$, $\phi_b = 158°$, (c) $\phi_a = 70°$, $\phi_b = 130°$, (d) $\phi_a = 60°$, $\phi_b = 160°$.

therefore necessary to modify the algorithm introduced in §6.3. Since some Q_q can be expected to lie outside $V_1V_2V_3V_4$, even when Q_1 is located correctly, a tolerance t is needed to permit acceptance of those Q_q which do not lie too far outside the parallelogram. If the perpendicular distance between any such Q_q and the side $V_1V_2V_3V_4$ closest to it is less than t then one says that this Q_q is acceptable (i.e. it is effectively located within the parallelogram).

For any Q_1 which leads to K acceptable perimeter points Q_q, the maximum number K_a of these Q_q are found through which a convex polygon can be constructed. This polygon is taken to be an estimate of $\sigma(\mathbf{x})$, whose interior represents an estimate of Υ_-, projections through which are computed at the angles of the given projections. The maximum difference Δ between the computed and given projections is calculated. This procedure is carried out for all Q which lead to K acceptable Q_q.

It is claimed that the best estimate of $\sigma(\mathbf{x})$, obtainable from the given projections, is that polygon for which Δ is least. Note from (6.8) that

$$J \approx (\ln K)/\ln N \tag{6.9}$$

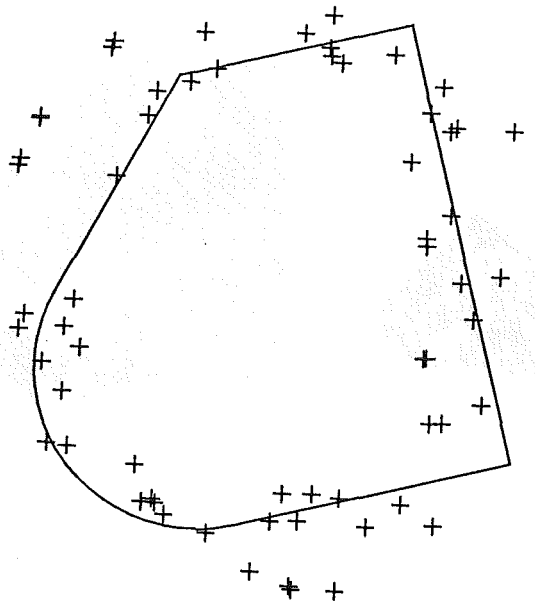
when is N is large. So, the number of projection widths that must be taken from each projection, in order to reconstruct any desired number K of perimeter points, decreases significantly with N, thereby appreciably lessening the wander of the Q_q.

Fig. 6.5a shows 65 perimeter points of the cross-section shown in fig. 6.4 reconstructed from the same pair of projections, but contaminated to a level of 5% (see §2.2.11). The perimeter points in fig. 6.5a wander from $\sigma(\mathbf{x})$ considerably more than they do in fig. 6.4. Fig. 6.5b shows the same number of perimeter points reconstructed from 8 projections, each contaminated to the level of 5%. The wander is seen to be much reduced, as is confirmed by the two values for Δ quoted in the caption to fig. 6.5.

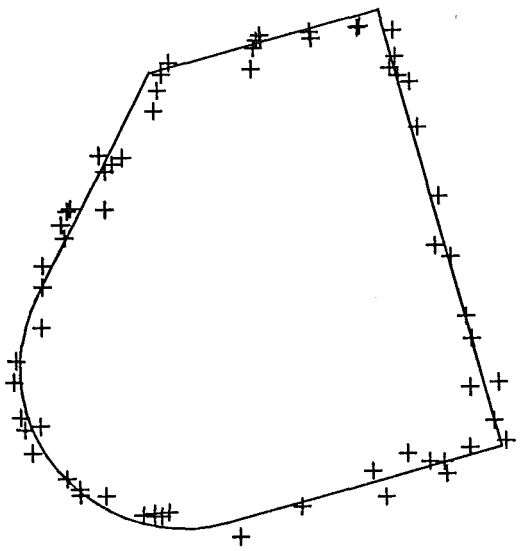
6.5 Preliminary Estimation of Perimeter of Non-Convex Cross-Section

Unaided, the algorithm introduced in §6.3 is impotent if $\sigma(\mathbf{x})$ possesses concavities or Υ_- is multiply connected. Some preliminary estimate of the shape of $\sigma(\mathbf{x})$ is required. To obtain this, the previous discovery, that boundaries of piecewise constant density regions of cross-sections can often be recognised in reconstructions from few projections, is invoked (Garden and Bates, 1984; Garden *et al.*, 1984). The following simple strategy for estimating $\sigma(\mathbf{x})$ is adopted.

First, the few given projections are subjected to modified (or filtered) back-projection (see §4.2.1.1), which is now the standard image reconstruction algorithm for computed (or computer-assisted) tomography (Bates and McDonnell, 1989, Chapter 5). Next, the shape of $\sigma(\mathbf{x})$, or the shapes of the separated parts of $\sigma(\mathbf{x})$ if Υ_- seems to be multiple-connected, are estimated by visual inspection. Finally, the estimate of this shape (or shapes) is recorded. Fig. 6.6 illustrates this. Fig. 6.6b shows the reconstruction by modified back-projection from 6 projections of the binary cross-section depicted in fig. 6.6a. Despite the small number of given projections, fig. 6.6b strongly suggests that Υ_- has no more than two separated parts. Fig. 6.6c shows estimate of $\sigma(\mathbf{x})$, obtained by visual inspection of fig. 6.6b, superimposed upon the true $\sigma(\mathbf{x})$.

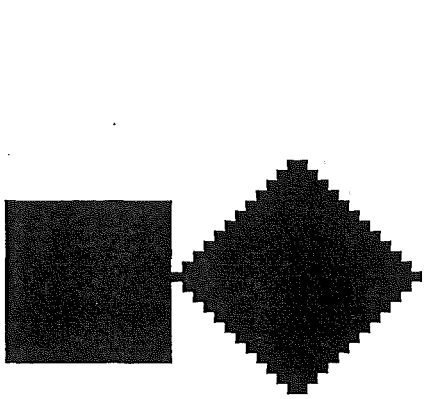


(a)

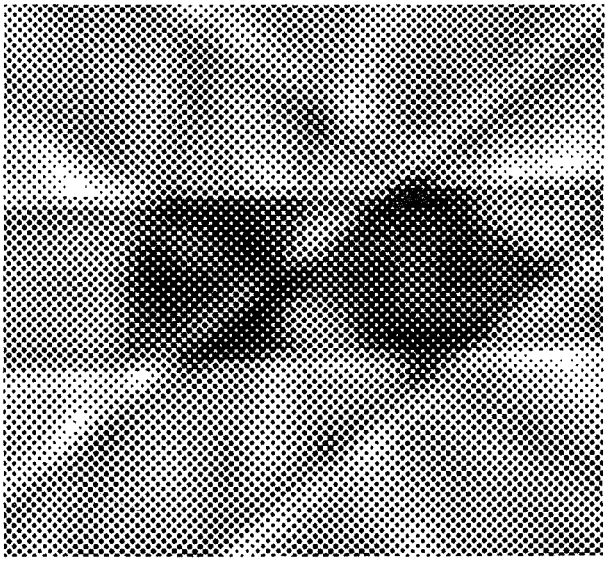


(b)

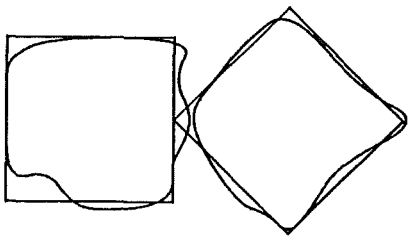
Figure 6.5: Convex cross-section reconstructed from contaminated ($\epsilon = 5$ % (see §2.2.11)) projections: (a) $N = 2$ (at angles 60° and 160°), $K = 64$, $t = 5$ pixels, $\Delta = 10$ pixels; (b) $N = 8$ (at angles 35°, 50°, 65°, 80°, 120°, 140°, 155°, 175°), $K = 64$, $t = 5$ pixels, $\Delta = 2$ pixels. The quoted values for Δ correspond to theses Q_1 which lead to minimum values of Δ (see §6.4). Full lines represent $\sigma(\mathbf{x})$. Crosses identify reconstructed perimeter points.



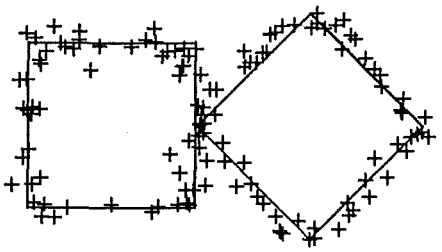
(a)



(b)



(c)



(d)

Figure 6.6: Non-convex cross-section reconstructed from 5 perfect projections, at angles 10°, 46°, 82°, 118°, and 154°. (a) true binary cross-section (black areas correspond to Υ_-); (b) reconstruction by modified back-projection; (c) estimate of $\sigma(\mathbf{x})$ (dashed lines) obtained by visual inspection of (b) superimposed upon true $\sigma(\mathbf{x})$ (full lines); (d) 144 Q_q, reconstructed by algorithm introduced in §7, superimposed upon $\sigma(\mathbf{x})$.

The extension of the above strategy is obvious when Υ_- is either single-connected and concave, or is multiple-connected with more than two parts, or is multiple-connected with more than two parts, or is multiple-connected with one or more of its parts being concave.

6.6 Reconstructing Non-Convex Cross-Sections from Few Projections

Having estimated the shape $\sigma(\mathbf{x})$ by visual inspection (as explained in §6.5), each projection width is partitioned in the manner described below and illustrated in fig. 6.7, wherein $\Upsilon_{-e,n}$ denotes the interior of the n^{th} of the estimated parts of $\sigma(\mathbf{x})$.

From the width theorem (§6.2) it is known that each projection width equals the sum of the widths, along the indicated line in the indicated direction, of all parts of Υ_-. So, along any line which (appears to) intersect only a single part of Υ_-, the ends of the projection width are positioned symmetrically with respect to the two points on the estimate of $\sigma(\mathbf{x})$ which intersect the line colinear with the projection width. These end points may of course lie either inside or outside the estimate of $\sigma(\mathbf{x})$ (they lie inside in the particular instance illustrated by $\overset{(1)}{\longleftrightarrow}$). When the line colinear with the projection width intersects two or more separated parts of Υ_-, fractions of the projection width are apportioned to the parts in proportion to their widths. The ends of these fractional projection widths are positioned in the same way as explained above with relation to $\overset{(1)}{\longleftrightarrow}$. Fig. 6.6d illustrates how this algorithm performs when presented with the 6 projections from which the image shown in fig. 6.6b was reconstructed. The reconstructed perimeter points are seen to cluster encouragingly closely around the true $\sigma(\mathbf{x})$.

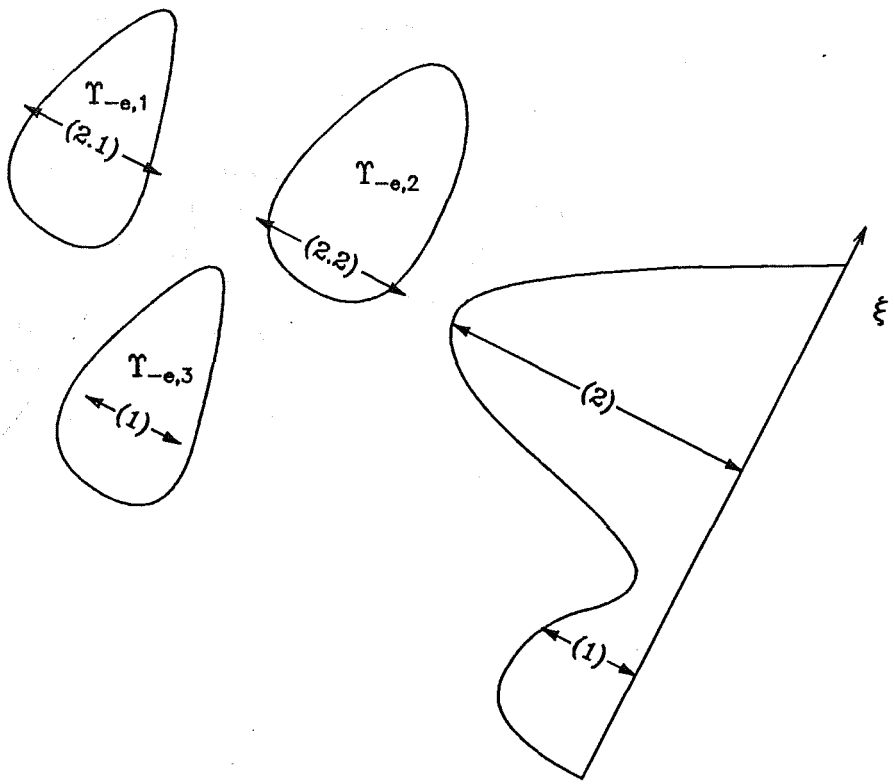


Figure 6.7: Apportioning projection widths between different parts of visual estimate of $\sigma(\mathbf{x})$, which is illustratively represented in the figure by the boundaries of the three regions $\Upsilon_{-e,1}$, $\Upsilon_{-e,2}$, $\Upsilon_{-e,3}$. The positioning of (parts of) each projection width is explained in §7. Note that the sum of the lengths $\overset{(2.1)}{\longleftrightarrow}$ and $\overset{(2.2)}{\longleftrightarrow}$ equals the length of $\overset{(2)}{\longleftrightarrow}$.

Chapter 7

Three Dimensional Object Orientation and Reconstruction using Silhouettes

In the past three decades the industrial world has seen a rapid growth in the use and application of robotics (Hungwen and Kender, Eds.) (1988). The modern robot however, while being a useful industrial tool, cannot operate in unfamiliar environments without supervision. The robot which can identify and pick up parts from a randomly filled bin, the universal bin picking robot, has not yet been invented. Robots able to identify objects in a completely uncontrolled environment, suitable for driving a vehicle say, have also not been developed. Central to the development of such robots are appropriate vision systems. Two main areas in which developments in robot vision are required are increased speed of the hardware and improved image analysis and processing (Braggins and Hollingum, 1986).

Introduced here is an original technique for image analysis. The orientation and shape of an unknown object in three-dimensional space are estimated from the object's silhouettes (Murch *et al.*, 1988b; McGregor, 1988; Murch and Bates, 1990b). The relative three-dimensional orientations of the silhouettes are initially unknown and have to be deduced before the object's shape can be reconstructed. Because silhouettes are binary images, practical implementation of the technique would not be particularly demanding of sensors, scene illumination, image processing requirements or system sensitivity to the environment.

Image analysis suitable for robot vision is reviewed in §7.1. Necessary notation is introduced in §7.2 as a preliminary to describing, in §7.3, the technique for estimating the orientation and shape of three-dimensional objects from their silhouettes. The algorithmic implementation of the technique is presented in §7.4, in which results of applying the algorithm to simulated and real-world data are reported. The significance of the technique is assessed in Chapter 8.

7.1 Image Analysis for Robotic Vision

A robot vision system can be understood in the context of its human counterpart. A "robot vision system" incorporates an "eye" (imaging instrument for viewing the scene), a "retina" (image capture device, e.g. photo-diode array), a "cortex" (computer for

operating on the captured images), a "memory" (data storage addressed by the "cortex" when interpreting captured images and relating them to previously stored information) and an "ear/mouth" (any input-output device(s) for communicating with other parts of the overall robot system or for transmitting processed information to, and receiving commands from, an interrogator). The memory must be partly "inherited" (comprising the software which controls the vision system and permits it to operate usefully on captured images) and partly "acquired" (consisting of captured information). The latter part of the memory is likely to be far larger than the former part.

The cortex corresponds (ideally) to the intelligence of the human vision system and performs the image analysis. Its task is to recognise objects and their movements in a scene. The first step towards performing this task is to extract features from the image. These features can then be invoked to model the scene. Matching of the scene model to known objects is the final step in object recognition. Once objects have been recognised, their movements can in principle be determined from examination of a sequence of images of the scene.

7.1.1 Feature Extraction

Some common features extracted from images are intensity, colour, edges, texture (Witkin, 1981), shading (Horn, 1975), centroids (Aggarwal and Martin, 1983), Fourier descriptors (Gardenier *et al.*, 1986) and range estimates (Rosenfeld, 1984).

Features can be classified as either global or local. The former are invariant to shift, rotation and scale whereas the latter are not. Centroids for example are global whereas edges are local.

Features are only useful if they serve to distinguish particular objects from others (Nagel, 1983). They should also be as insensitive as possible to noise and scene illumination (Poggio *et al.*, 1985). Furthermore, their extraction and subsequent processing should not be excessively expensive.

7.1.2 Scene Models

Once features of an image have been extracted they must be efficiently organised into a model of the scene. Such a model is needed for interpreting the scene. The four types of scene models, which are commonly invoked, are often referred to as pixelated, segmented, geometrical and relational (Ballard and Brown, 1982).

A pixelated model is the captured image itself. Intensity or colour of the image's pixels are associated directly with the pixelated model of the scene. Pixelated models were used in the earliest of image analyses (Ballard and Brown, 1982).

Segmented models are based on identifying separate regions in images where common features are to be found (Brady, 1982). A particular object is then, hopefully, located in each such region. For example, an image may be separated into regions of similar texture so that each region corresponds to an object in the scene.

Geometrical models are based on the two- and three-dimensional shapes of objects which are present in scenes (Kemper and Wallrath, 1987). Edges or shades of an image are used to create three-dimensional shapes or outlines of objects (Richards *et al.*, 1987).

Relational models are extensions of segmented models (Rosenfeld, 1984). Information present in a segmented model is represented by a graph in which the nodes correspond to regions or features. The arcs joining the nodes have associated with

them relational values. Relational models are useful for high level processing of images (Ballard and Brown, 1982).

7.1.3 Object Recognition

Once the scene has been appropriately modelled, the individual objects must be recognised. To achieve this, catalogues of known objects are compared with objects in the scene model. Such a catalogue must contain models of all the objects one wishes to recognise. The comparison, or matching, process which is most appropriate usually depends on the particular model and features that are adopted (Rosenfeld, 1988; Chin and Dyer, 1986). Recognition should ideally be independent of the directions from which objects are viewed, no matter what are the backgrounds or the conditions of illumination. Object recognition is a part of the huge field of pattern recognition (Duda and Hart, 1973; Fu, 1982).

Recognition schemes can be based on the pixelated model (Ballard and Brown, 1982). An object is recognised by correlating the model with the catalogue of object models. Any objects in the catalogue which are sufficiently well correlated are said to be recognised.

One disadvantage of such techniques stems from the scene model being essentially two-dimensional, while the scene and the objects in it are generally three-dimensional (Rosenfeld, 1984). In order to be able to recognise an object, many views of it must be contained in the object model. Furthermore, each of these views must be correlated against objects in the scene before the object can be said to have been recognised. Consequently, features and scene models which incorporate the underlying three-dimensional nature of scenes are most desirable. Features, such as shading and occlusion, used in geometrical and relational scene models, are obviously more suitable for this than the features which are commonly incorporated into pixelated models (Connolly *et al.*, 1987). An approach which might be of some use here could be based on the orientation and shape estimation technique presented in §7.2 through §7.4. A popular matching technique for geometrical models is parameter optimisation (Grimson, 1987), for which the distances between parameters representing the object in the scene and the object models are minimised. Matching of relational models is performed by finding isomorphisms between unknown objects in the scene and the catalogue of objects (Ballard and Brown, 1982).

The recognition problem has many facets which have suggested a host of possible solutions. It seems that one cannot hope to be successful without employing a large number of different features (Sanz, 1988; Sanz and Hinkle, 1988). Other useful approaches are based on expert or knowledge-based systems (Kak *et al.*, 1988), perceptual models (Lowe, 1987) and neural networks (Fukushima, 1988).

7.1.4 Motion Estimation

Any effective solution to the object recognition problem should make it possible for robots to generate sequences of faithful static models of observed scenes. For a robot to be capable of performing economically useful tasks, without continual remote supervision, it is necessary for the vision system's cortex to infer the dynamics as well as the statics of both the robot and whatever objects are within its field of view. Central to this is estimating relative orientation and linear translation from sequences of images.

A variety of approaches to collating and processing sequences of images has been collected by Huang (1981). The versatility and wide applicability of the Erlangen-bispectrum should be noted (Bates, 1987).

In §7.2 through §7.4 results from an original approach to solving the orientation estimation problem, by operating upon an object's silhouettes, is reported. Related approaches have of course been put forward previously (Wang *et al.*, 1984; Potmesil, 1987; Beusmans *et al.*, 1987; Ahuja and Veenstra, 1989), but the inspiration for this technique is different in that it stems from the standard theory of image reconstruction from projections (Barrett, 1981; Bates and McDonnell, 1989, Chapter 5). Other approaches, which complement methods which employ silhouettes, are those based on extremum principles (Brady and Yuille, 1984; Weiss, 1988) and simplified mathematical descriptions of surfaces (Ikeuchi and Horn, 1981; Richards *et al.*, 1987).

7.2 Notational Preliminaries

Arbitrarily chosen (but considered fixed, once chosen) Cartesian coordinates x, y and z, which can be thought of as components of the position vector $\mathbf{x}$, are set up in three-dimensional image space (sometimes called configuration or object space). A solid, or opaque, object $\chi(\mathbf{x})$ can then by defined by

$$\chi(\mathbf{x}) = \begin{cases} 1 & \text{for} \quad \mathbf{x} \in \Upsilon_- \\ 0 & \text{for} \quad \mathbf{x} \in \Upsilon_+ \end{cases} \tag{7.1}$$

where the terminology defined in §1.3.1 is invoked. It is consistent with normal usage to refer to the surface $\sigma(\mathbf{x})$ of the object as the object's shape. Because only the shape of the object is of interest, no generality is lost by defining $\chi(\mathbf{x}) = 1$ in Υ_-.

The three-dimensional Fourier transform of $\chi(\mathbf{x})$ is denoted $\tilde{\chi}(\mathbf{u})$, with $\mathbf{u}$ being the position vector of an arbitrary point in three-dimensional Fourier space whose Cartesian components are u, v and w. An arbitrary point in the plane in Fourier space complementary to the x, y-plane in image space is $(u, v, 0)$. When $\mathbf{u}$ is specialised to $(u, v, 0)$, $\tilde{\chi}(\mathbf{u})$ is written as $\tilde{\chi}(u, v)$.

The notation $\chi_{\alpha,\beta,\gamma}(\mathbf{x})$ implies that the object is rotated through the Euler angles α, β, γ (Arfken, 1985, Chapter 4). The use of the Euler angles for specifying object rotation is illustrated in fig. 7.1b.

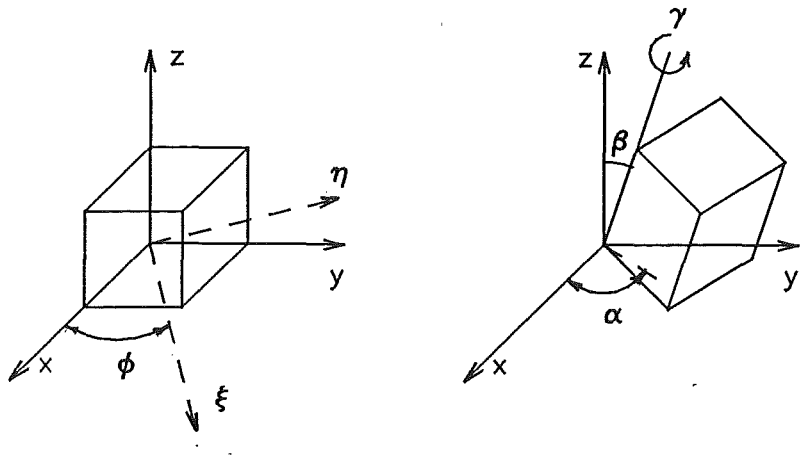
The two-dimensional projection $p(x, y)$ of $\chi(\mathbf{x})$ is defined by line integrals of the object parallel to the z-axis. This means that $p(x, y)$ is identical to $p(\xi; \eta; \theta)$, as defined by (4.6). On replacing ζ in (4.6) by z, it is seen that

$$p(x, y) = \int_{-\infty}^{\infty} \chi(\mathbf{x})\, dz \tag{7.2}$$

The one-dimensional projection $p(\xi; \phi)$ of $\chi(\mathbf{x})$ at angle ϕ, which is defined by (4.9), is given in terms of $p(x, y)$ by

$$p(\xi; \phi) = \int_{-\infty}^{\infty} p(x, y)\, d\eta \tag{7.3}$$

where the (ξ, η) coordinates have the same origin as, but are rotated through ϕ with respect to, the (x, y) coordinates, as is illustrated in fig. 7.1a. A one-dimensional projection is typically displayed as a curve, e.g. showing how $p(\xi; \phi)$ varies with ξ for a



(a) (b)

Figure 7.1: Coordinate systems specifying object rotations: (a) coordinates for one-dimensional projections of silhouettes; (b) Euler angles α, β, γ (all points in the object are rotated by α, β and γ as indicated).

particular value of ϕ. The projection can also be represented as a straight line (i.e. the ξ axis) of varying brightness. This brightness is proportional to $p(\xi;\phi)$. When equi-spaced (in ϕ) projections are displayed side by side in this manner, a two-dimensional image is formed. This image, which is appropriately denoted $p(\xi;\phi)$, but with both ξ and ϕ being variable, is known as a sinogram (Herman, 1980, Chapter 1). The term sinogram arises because high contrast points in $p(x,y)$ give rise to distinctive sinusoidal curves in $p(\xi;\phi)$.

The silhouette $s(x,y)$ of the object, viewed in the z-direction, is defined by

$$s(x,y) = \begin{cases} 1 & \text{if} \quad p(x,y) > 0 \\ 0 & \text{if} \quad p(x,y) = 0 \end{cases} \tag{7.4}$$

It is convenient to denote the one-dimensional projection, at angle ϕ, of the silhouette viewed in the z-direction, by $ps(\xi;\phi)$ which is defined by the integral in (7.3) but with $p(x,y)$ replaced by $s(x,y)$.

The positivity operator $\mathcal{P}$ is here defined, with respect to an arbitrary real function $h(\xi;\phi)$, by

$$\mathcal{P}h(\xi,\phi) = \begin{cases} 1 & \text{if} \quad h(\xi,\phi) > 0 \\ 0 & \text{if} \quad h(\xi,\phi) = 0 \end{cases} \tag{7.5}$$

It follows from the above definitions, that

$$\mathcal{P}ps(\xi;\phi) = \mathcal{P}p(\xi;\phi) \tag{7.6}$$

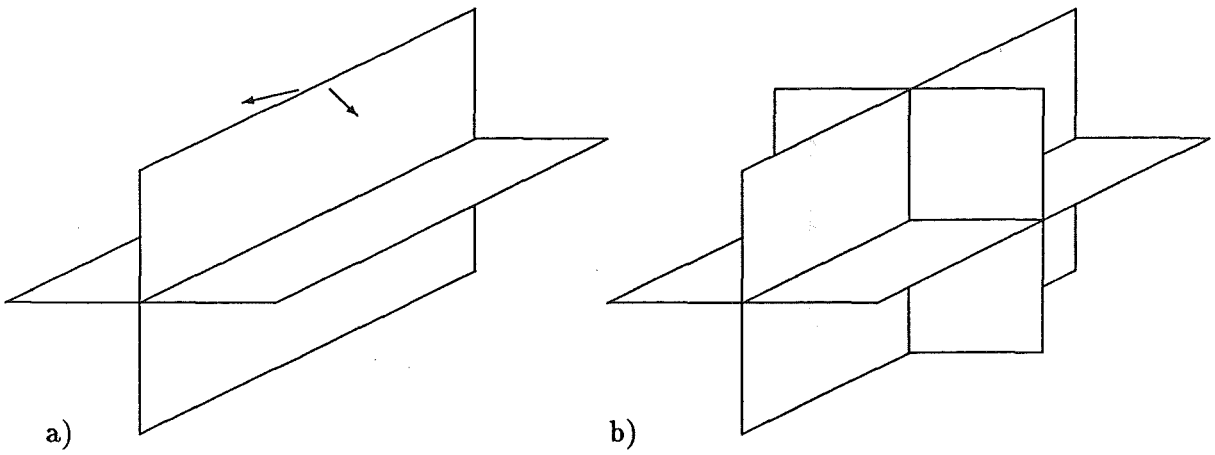


Figure 7.2: The intersection of slices in three-dimensional Fourier space. (a) Two slices in Fourier space (the slices intersect along a common Fourier radial; the two projections can however be rotated about this common radial); (b) the inclusion of a third slice fixes the actual orientations of all slices.

When the $\mathcal{P}ps(\xi;\phi)$, for equi-spaced values of ϕ, are displayed side by side as brightness-modulated lines, the resultant two-dimensional image is here called a binary sinogram. Since $\mathcal{P}ps(\xi;\phi)$ only assumes the values zero or one, only two brightness levels are possible. Here, the brightness levels white and black respectively correspond to zero and one.

7.3 Estimating Object Orientation and Shape from Silhouettes

Although the concern here is with silhouettes, it is convenient to begin by considering projections and their associated theorems.

The three-dimensional manifestation of the projection theorem (see §4.2.1.1) states that $\tilde{\chi}(u,v)$ is the two-dimensional Fourier transform of $p(x,y)$. This can be expressed loosely, but graphically, as the Fourier transform of $p(x,y)$ being a slice through $\tilde{\chi}(\mathbf{u})$, along the u,v-plane in Fourier space.

Because $\chi_{\alpha,\beta,\gamma}(\mathbf{x})$ is merely a rotated version of $\chi(\mathbf{x})$, the Fourier transform of $p_{\alpha,\beta,\gamma}(x,y)$ also represents a slice through $\tilde{\chi}(\mathbf{u})$ but rotated with respect to the previous slice by the Euler angles α,β,γ. The intersection of these two slices in Fourier space is here called the common Fourier radial. This is illustrated in fig. 7.2a.

The two-dimensional manifestation of the projection theorem (see §4.2.1.1) states that the common Fourier radial is the Fourier transform of the one-dimensional projection of $p(x,y)$ taken at an angle corresponding to the normal to the line of intersection. It follows that, if the one-dimensional projection common to both two-dimensional projections is found, the orientation of the common Fourier radial can be determined. This is expressed as

$$p(\xi;\phi) = p_{\alpha,\beta,\gamma}(\xi;\phi') \tag{7.7}$$

When (7.7) is satisfied, the angles ϕ and ϕ' determine the orientation of the common

Fourier radial.

Given only two projections, however, their actual orientations cannot be uniquely inferred. This is because rotation is possible about the Fourier radial, as illustrated in fig. 7.2a. The addition of a third two-dimensional projection $p_{\alpha',\beta',\gamma'}(x,y)$ eliminates this non-uniqueness by introducing two further common Fourier radials. This is illustrated in fig. 7.2b. It follows that the actual orientation of the object, as viewed in the directions at which the original projections were taken, can be determined (van Heel, 1987).

Since most objects are opaque, a robot's vision system cannot measure their projections, but it can record their silhouettes. Because of (7.6) and (7.7), it is appropriate to collect pairs of one-dimensional projections of silhouettes for which

$$\mathcal{P}ps_{\alpha_l,\beta_l,\gamma_l}(\xi;\phi) = \mathcal{P}ps_{\alpha_m,\beta_m,\gamma_m}(\xi;\phi') \tag{7.8}$$

with $(\phi,\alpha_l,\beta_l,\gamma_l)$ differing from $(\phi',\alpha_m,\beta_m,\gamma_m)$ for all $l \neq m$. Although each such pair does not necessarily determine the orientation of the common Fourier radial, one pair is guaranteed to by (7.6). I refer to this as the Fourier radial ambiguity.

Given two silhouettes, characterised by the two sets of angles $\alpha_l,\beta_l,\gamma_l$ and $\alpha_m,\beta_m,\gamma_m$ a pair ϕ,ϕ' of angles satisfying (7.8) are found. Note, first, that ϕ,ϕ' may not correctly identify an actual common Fourier radial. Note further that, as remarked above, even if this is the location of a common Fourier radial, the orientation of one silhouette relative to the other is not fixed, because either silhouette can be rotated about the common Fourier radial without contravening (7.8). Given another silhouette, however, and assuming that the common Fourier radials for all three pairs of silhouettes are correctly identified, there is no longer any rotational freedom, so that the relative orientations of the three silhouettes are fixed. Algebraic manipulation (van Heel, 1987) readily permits this procedure to be transcribed into a computational algorithm, which generates what I call an elementary skeleton of the object (i.e. the three silhouettes fitted together). The skeleton is an approximation to the convex hull of the object. The term convex shape is adopted to denote the approximation to the convex hull since it is both graphic and intuitively satisfying.

Because of the aforementioned Fourier radial ambiguity, the first attempt at constructing an elementary skeleton may not be successful. To check this, the silhouettes of the elementary skeleton are constructed. The skeleton is consistent only if these silhouettes are identical to the original ones. If, however, one of the silhouettes which make up the skeleton is wrongly positioned, it protrudes beyond one or both of the other two silhouettes when the skeleton is viewed in the directions corresponding to the latter pair of silhouettes. The offending silhouette must be repositioned, but in such a way as to satisfy (7.8) for both of the other silhouettes, so that it no longer protrudes. This procedure is repeated in turn for each of the three silhouettes until the elementary skeleton is found to be consistent. I call this process the self consistency check. Given more than three silhouettes, they can be manipulated in triples as described above. This is illustrated by example in §7.4.

Before this approach can be implemented, the silhouettes must be registered (i.e. referred to some identifiable point in space). To perform this it is necessary to establish what is appropriately called a chief reference point (van Dyck, 1987). A corner is an ideal reference point. Whatever reference point is chosen, it is necessarily obscured in some silhouettes, because of the object's opaqueness. Ancillary reference points must

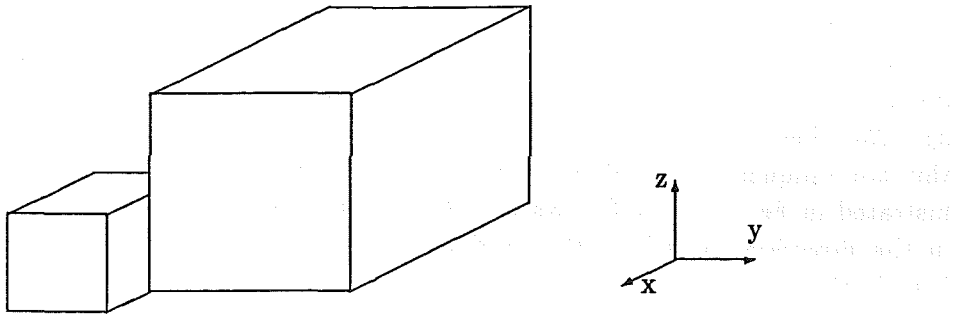


Figure 7.3: An object consisting of two contiguous solid cubes. The reference coordinate system is included in this figure to help the reader understand the specification, in the first paragraph of §7.4, of the orientations of this object's silhouettes.

be chosen in those silhouettes in which the chief reference point is obscured. Each ancillary reference point must be visible together with the chief reference point in some silhouettes, thereby allowing the (sometimes hidden) position of the chief reference point to be transformed to all silhouettes.

A question which naturally arises from the self consistency check is uniqueness. Given three silhouettes, is there only one object which is self consistent with them? Because a two-dimensional projection contains all the information that its corresponding silhouette does, it follows that, if three silhouettes are not self consistent, then neither can the three projections be self consistent. To show that projections are self consistent, it is necessary to establish that the actual orientation and shape of the object can be uniquely determined from three of its two-dimensional projections. The actual orientations of the three projections can be found in the manner described earlier in this section. The motivation for the development of the algorithm presented in Chapter 6 was to show that three projections are sufficient to determine the shape of an object. In Chapter 6 it is demonstrated that only two one-dimensional projections are required to determine the shape of a convex cross-section. It consequently follows that three unoriented two-dimensional projections of a convex object are uniquely self consistent. However, it is still an open question whether this holds for non-convex objects.

7.4 Algorithmic Implementation

The algorithmic implementation of the technique presented in §7.3 is most readily understood by example. Consider the object shown in fig. 7.3. It comprises two contiguous solid cubes of different sizes. With respect to the coordinate system pictured in fig. 7.3, three silhouettes of the object, denoted l, m and n, are viewed at orientations of the object specified by $\alpha_l = 0°, \beta_l = -90°, \gamma_l = 0°$, $\alpha_m = 0°, \beta_m = -90°, \gamma_m = 90°$ and $\alpha_n = 0°, \beta_n = 0°, \gamma_n = 0°$. These silhouettes are depicted in fig. 7.4. It is shown below how these silhouettes, which are chosen here for ease of visualisation, can be used to reconstitute the orientations of the silhouettes with respect to the object.

To start the orientation estimation process, orientations which satisfy (7.8) must be found. For this purpose, the $\mathcal{P}ps(\xi;\phi)$ are calculated for each silhouette and are displayed as a binary sinogram. That is, the $\mathcal{P}ps(\xi;\phi)$, for equi-spaced values of ϕ, are displayed side by side as white and black (corresponding to the values zero and one

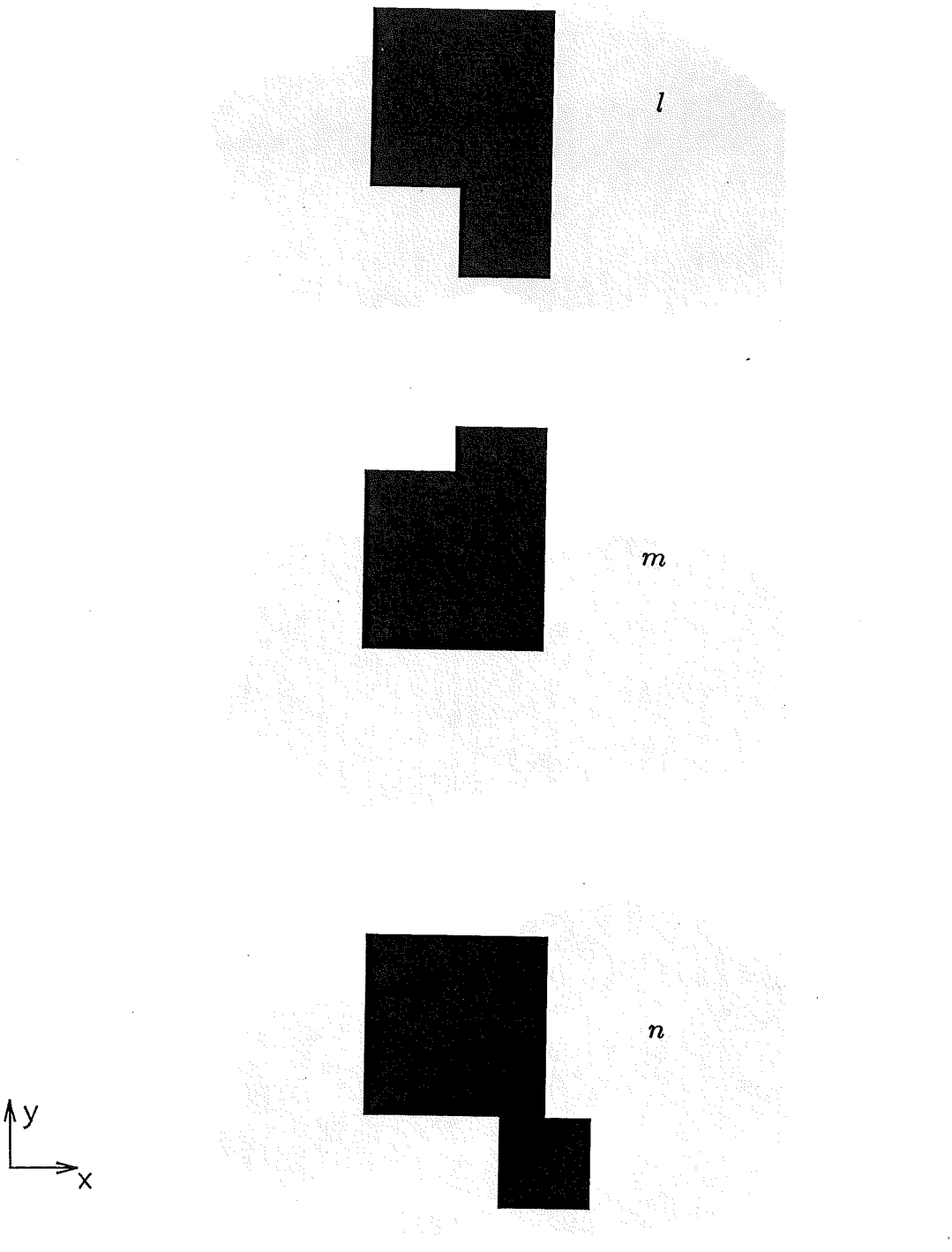


Figure 7.4: Silhouettes of the object illustrated in fig. 7.3. The top, middle and bottom silhouettes correspond, respectively, to the object being viewed at the orientations specified by $\alpha_l = 0°, \beta_l = -90°, \gamma_l = 0°$, $\alpha_m = 0°, \beta_m = -90°, \gamma_m = 90°$ and $\alpha_n = 0°, \beta_n = 0°, \gamma_n = 0°$.

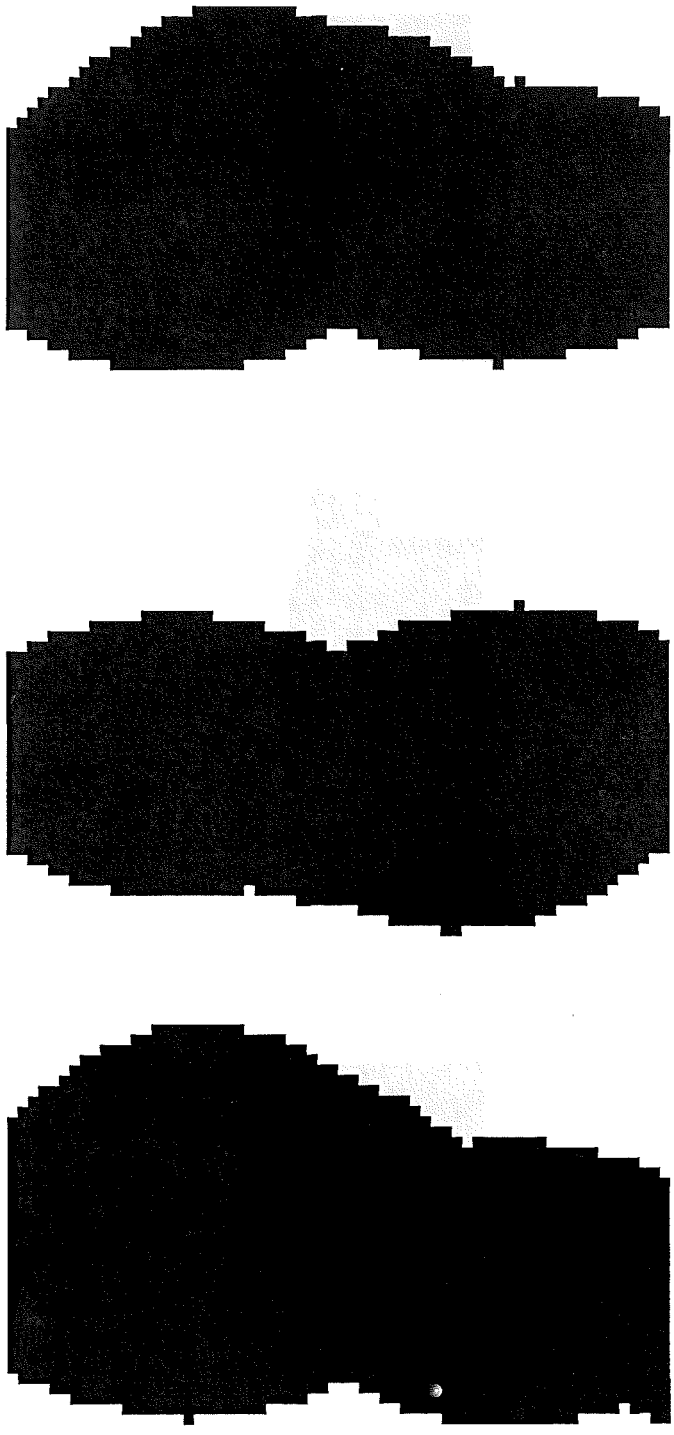


Figure 7.5: Binary sinograms of the silhouettes shown in fig. 7.4. The top, middle and bottom binary sinograms are, respectively, denoted l, m and n and correspond to binary sinograms of the l, m and n silhouettes shown in fig 7.4.

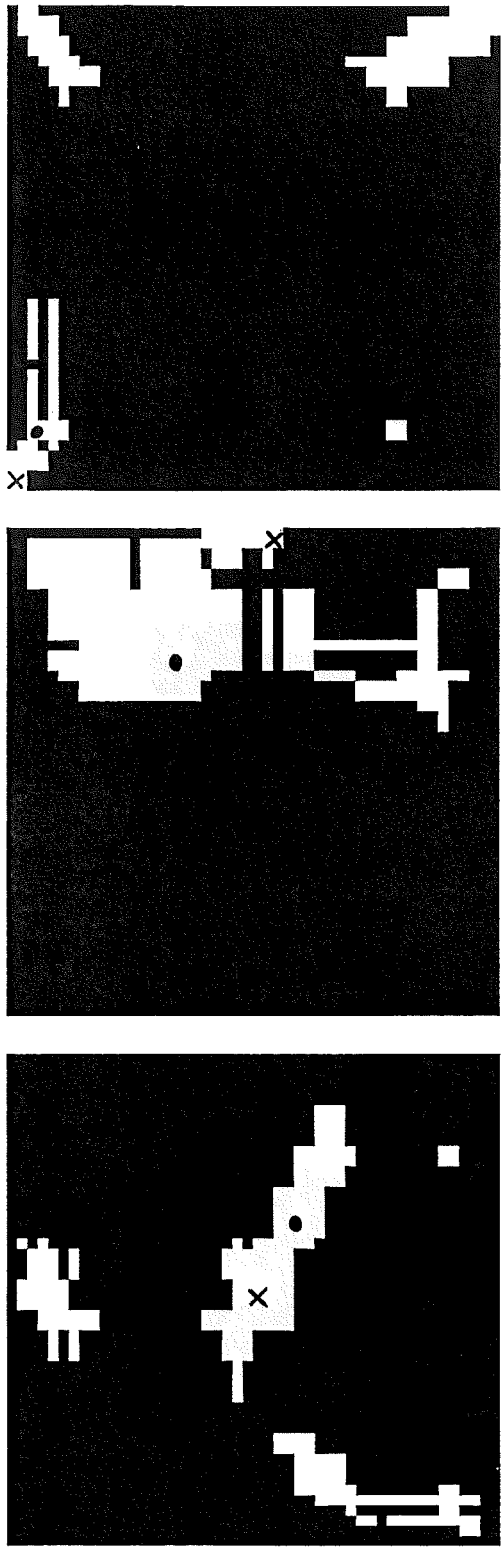


Figure 7.6: Images of the thresholded sccf functions (accepted points in white) resulting from the silhouettes shown in fig. 7.4. Respectively the top, middle and bottom sccf images correspond to the invocation of (7.9) for the pairs l, m and m, n and n, l of sinograms shown in fig. 7.5. Crosses on the white regions identify the actual object orientations whilst the points marked • correspond to a trial orientation.

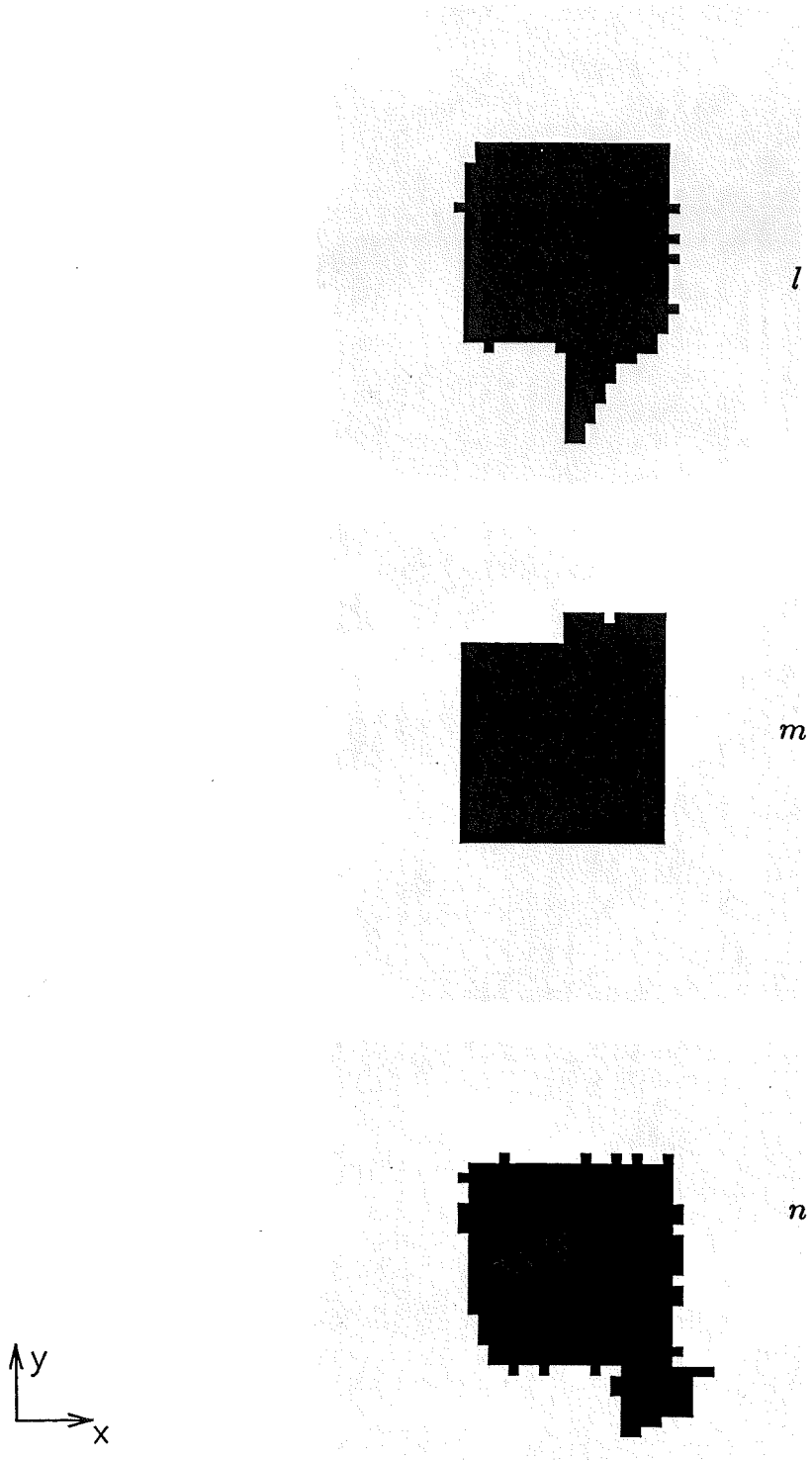


Figure 7.7: Silhouettes of an object generated by the self consistency check part of the algorithm presented in §7.3. The shape of the object is obtained by backprojecting the silhouettes shown in fig. 7.4 at the orientations specified by the points marked • in fig. 7.6.

respectively) modulated lines (see §7.2). Binary sinograms formed from the silhouettes shown in fig. 7.4 are displayed in fig. 7.5.

To find the orientations at which (7.8) is satisfied the binary sinograms are compared in the following manner. For each ϕ and ϕ' the quantity

$$\int_{-\infty}^{\infty} |\mathcal{P}ps_{\alpha_l,\beta_l,\gamma_l}(\xi;\phi) - \mathcal{P}ps_{\alpha_m,\beta_m,\gamma_m}(\xi;\phi')|\, d\xi = \mathrm{sccf}(\phi,\phi') \tag{7.9}$$

is calculated. The function $\mathrm{sccf}(\phi,\phi')$ is here termed the sinogram cross coefficient function (sccf) and its value is used as an estimate of the accuracy with which (7.8) is satisfied for each ϕ and ϕ'. When $\mathcal{P}ps_{\alpha_l,\beta_l,\gamma_l}(\xi;\phi)$ and $\mathcal{P}ps_{\alpha_m,\beta_m,\gamma_m}(\xi;\phi')$ are equal for a particular ϕ and ϕ', $\mathrm{sccf}(\phi,\phi')$ is zero. The sccf function can also be displayed as an image. Each point in the image represents the value of the sccf function for a particular ϕ and ϕ'. To find the ϕ and ϕ' that best satisfy (7.8), the sccf function is thresholded. That is, function values above an acceptance level are set to 1 whilst function values below this value are set to 0. The thresholded sccf functions of the binary sinograms shown in fig. 7.4 are pictured in fig. 7.6. The white areas correspond to values ϕ and ϕ' which satisfy (7.8). Each such ϕ, ϕ' pair is referred to as an accepted point.

An accepted point in each sccf function represents a possible orientation of the object at which each silhouette was viewed. Clearly there are many orientations at which (7.8) is satisfied. To estimate the correct orientation from this reduced set, the self consistency check is invoked. For the self consistency check, one accepted point is selected from each sccf function and the corresponding relative orientations are computed by simple vector manipulation (van Heel, 1987). Using these orientations, an approximation to the object's convex shape is reconstructed by backprojecting each of the silhouettes along rays normal to their respective planes. An estimate of the object's shape or skeleton is obtained where all three backprojections intersect. Silhouettes of this estimate of the convex shape are then computed and compared to the originals. If they do not match the original silhouettes, then it can be said that the correct orientation is not yet found. The remaining accepted points in the sccf functions are then similarly checked for self consistency, one after the other, until a set of points is found which satisfies self consistency.

Fig. 7.7 illustrates the silhouettes generated after the self consistency check is invoked for the points marked • in fig. 7.6. It is clear that the accepted points marked • in fig. 7.6 are not correctly oriented because the silhouettes are different from the originals.

If the points marked × are invoked in fig. 7.6 for the self consistency check then silhouettes are obtained which are visually indistinguishable from the originals shown in fig. 7.4. It is only for the points marked × that the self consistency check is satisfied. Hence the actual orientation of the object is recoverable uniquely from the points marked ×.

The algorithm is further illustrated with three silhouettes of an actual machined part, whose plan, elevation and end elevation are shown in fig. 7.8. Three silhouettes of this object are shown in fig. 7.9. With respect to the coordinate system indicated in fig. 7.8 the silhouettes are viewed at the object orientations specified by $\alpha_l = -55°, \beta_l = -90°, \gamma_l = 0°$, $\alpha_m = -45°, \beta_m = -90°, \gamma_m = -45°$ and $\alpha_n = 0°, \beta_n = -35°, \gamma_n = 90°$. In order to simulate the resolution typical of an industrial electronic camera (e.g. the machined part could be on a conveyor belt which is viewed by a robot's vision system),

the silhouettes shown in fig. 7.9 have been chosen so as to exhibit significantly less detail than is apparent in fig. 7.8. The images of the sccf functions are displayed in fig. 7.10. The crosses on these images identify the orientation at which the elementary skeleton (formed from the three silhouettes) is found to be consistent. The silhouettes obtained by invoking the points marked • in fig. 7.10 points are shown in fig. 7.11. Again it is only for the correct orientations, corresponding to the points marked ×, that the self consistency check is satisfied, further verifying this algorithm.

The algorithm is further illustrated with four silhouettes of a simulated object, whose plan, elevation and end elevation are shown in fig. 7.12. The four silhouettes of the object are shown in fig. 7.13. With respect to the coordinate system indicated in fig. 7.12 the silhouettes are viewed at the object orientations specified by $\alpha_l = -25°, \beta_l = -90°, \gamma_l = 0°$, $\alpha_m = -125°, \beta_m = -90°, \gamma_m = 46°$, $\alpha_n = 310°, \beta_n = -63°, \gamma_n = 159°$ and $\alpha_o = 0°, \beta_o = -35°, \gamma_o = 90°$. The binary sinograms corresponding to the l and m silhouettes are depicted in fig. 7.14. The six sccf functions are shown in fig. 7.15. As stated in §7.3, the actual orientations of the object for each silhouette can be determined by processing the silhouettes in triples. For four silhouettes, there are three triples. Only two triples, however, are required to determine the actual orientation of the object. The third triple can be used as a check. An example of silhouettes obtained from the points marked • in fig. 7.15 are presented in fig. 7.16. Again it is obvious that these points are not correct. When the points marked × are invoked, silhouettes which are visually indistinguishable from the originals shown in fig. 7.13 are obtained. Once again it is only for the correct orientations that the self consistency check is satisfied, providing additional confirmation of the correctness of the reconstruction algorithm.

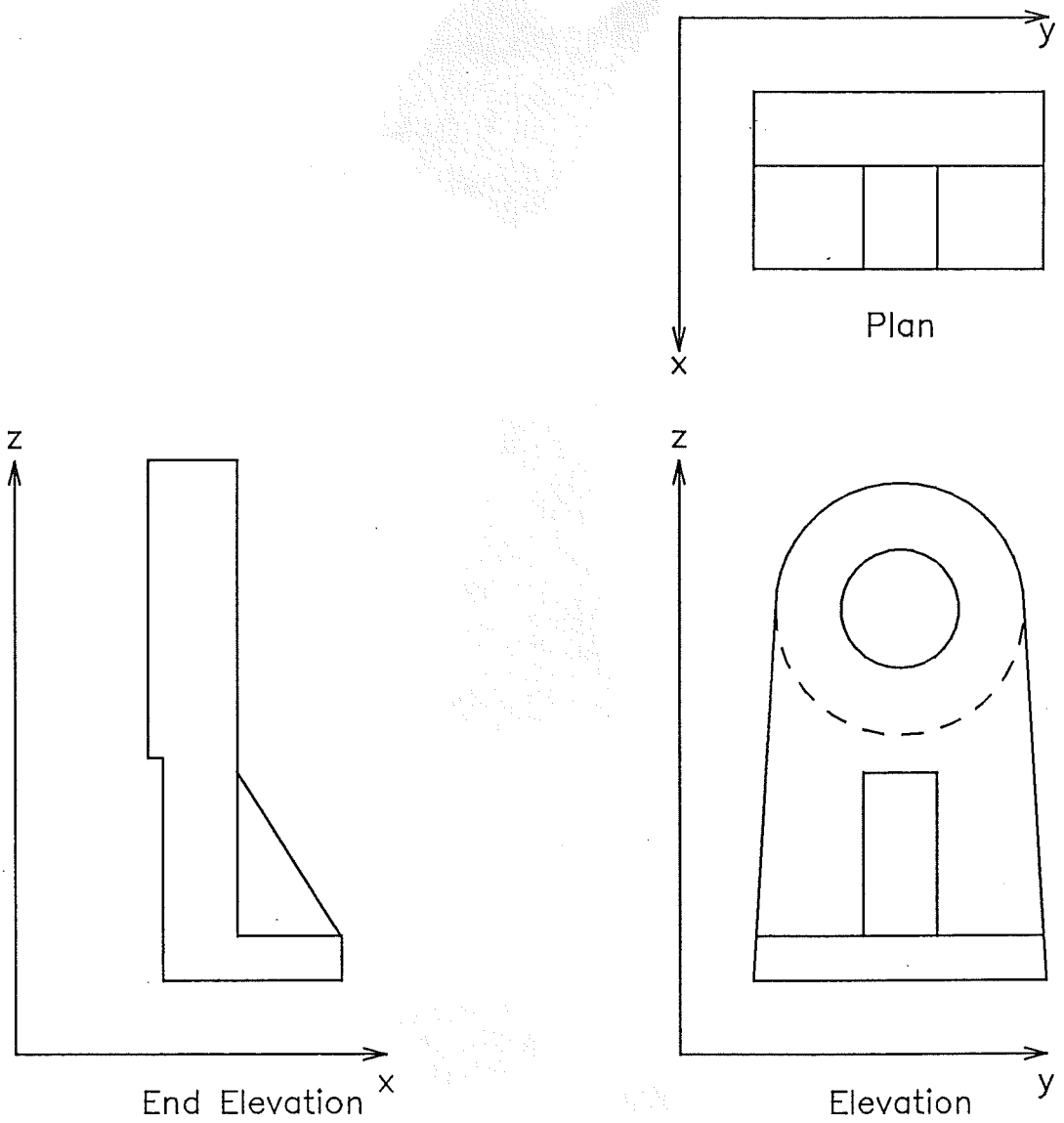


Figure 7.8: Orthographic drawing of a machined part.

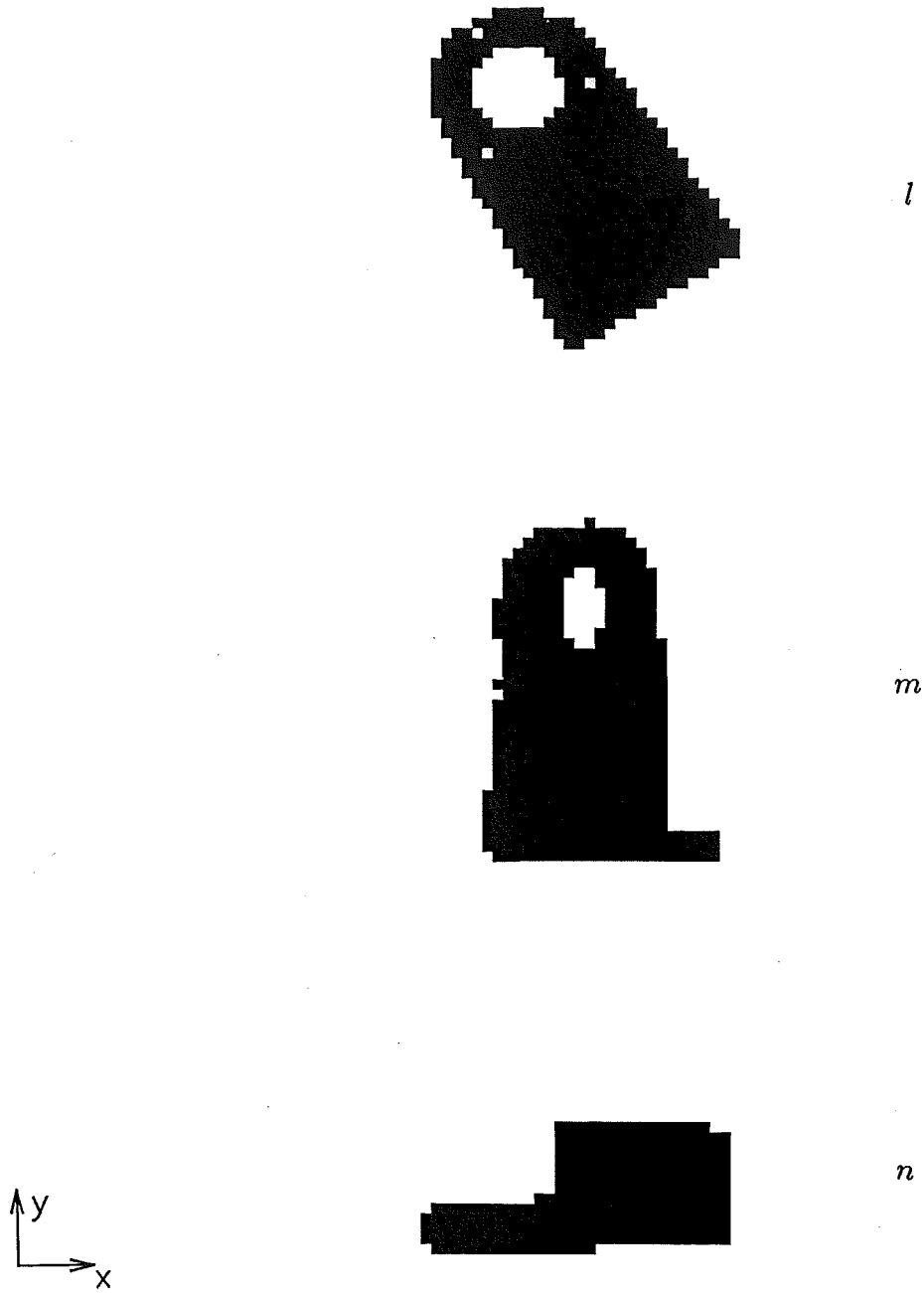


Figure 7.9: Three silhouettes of the machined part shown in fig. 7.8. The orientations at which this object was viewed, with respect to the coordinate system in fig. 7.8, was respectively $\alpha_l = -55°, \beta_l = -90°, \gamma_l = 0°$, $\alpha_m = -45°, \beta_m = -90°, \gamma_m = -45°$ and $\alpha_n = 0°, \beta_n = -35°, \gamma_n = 90°$.

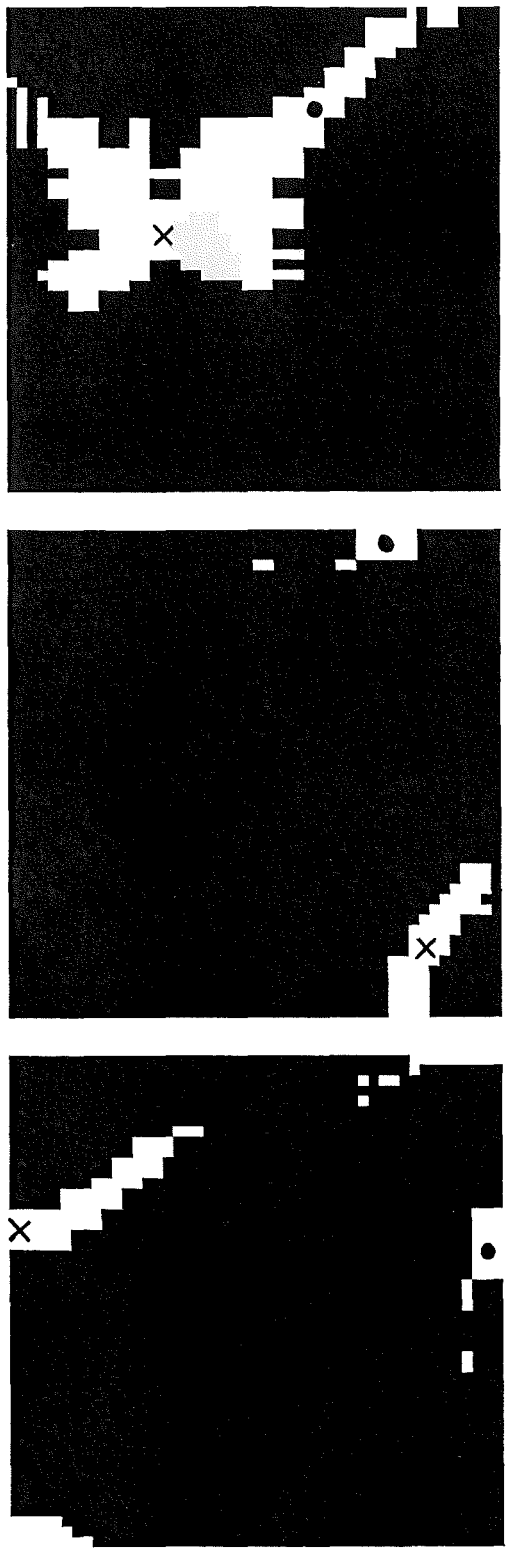


Figure 7.10: Images of the thresholded sccf functions (accepted points in white) resulting from the silhouettes shown in fig. 7.9. Respectively the top, middle and bottom sccf images correspond to the invocation of (7.9) for the pairs l, m and m, n and n, l of sinograms computed from the silhouettes shown in fig. 7.9. Crosses on the white regions identify the actual object orientations whilst the points marked • correspond to a trial orientation.

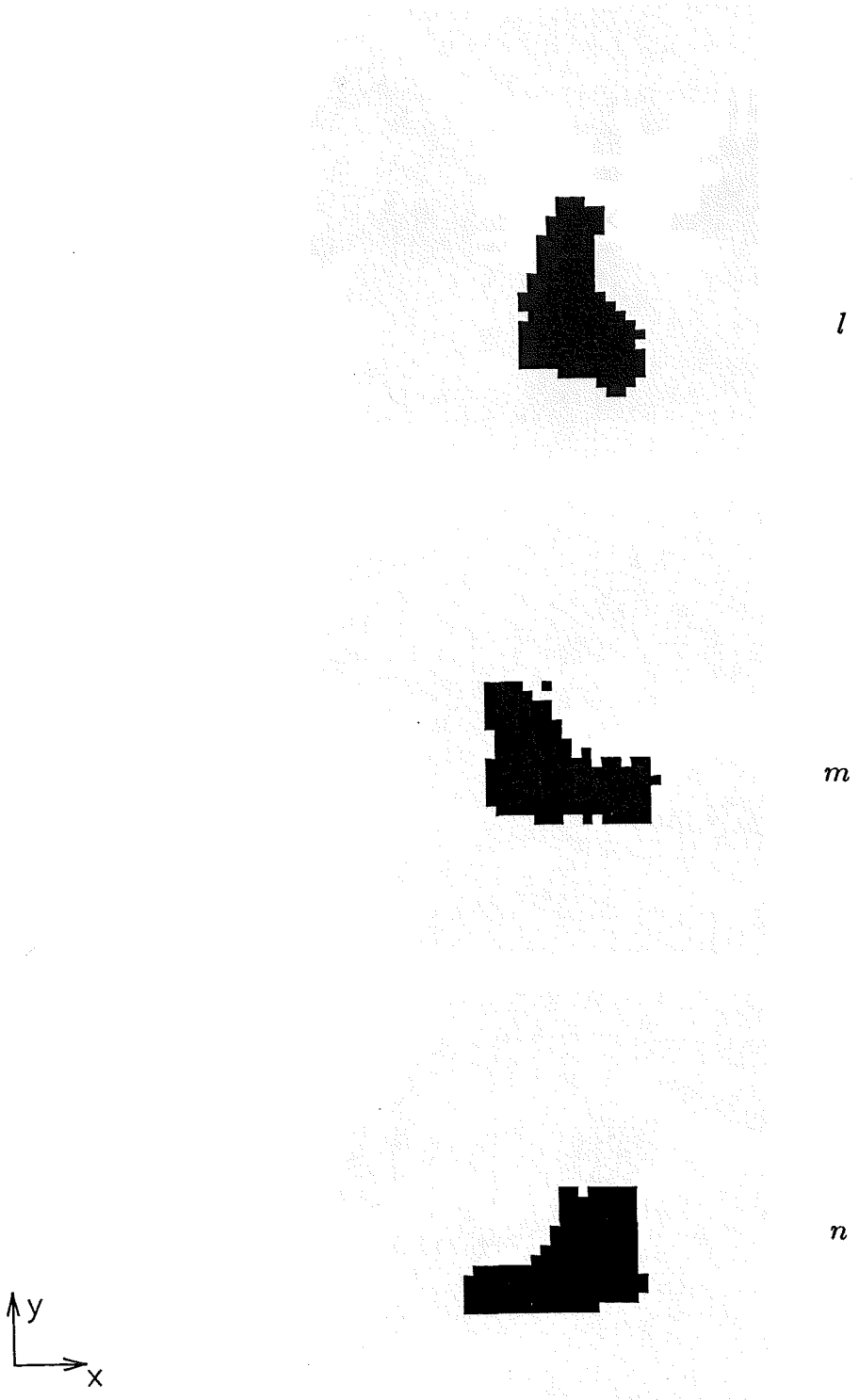


Figure 7.11: Reconstructed silhouettes of the machined part obtained by invoking the self consistency check incorporated into the algorithm presented in §7.3. The objects shape is obtained by backprojecting the silhouettes shown in fig. 7.9 at the orientations specified by the points marked • in fig. 7.10.

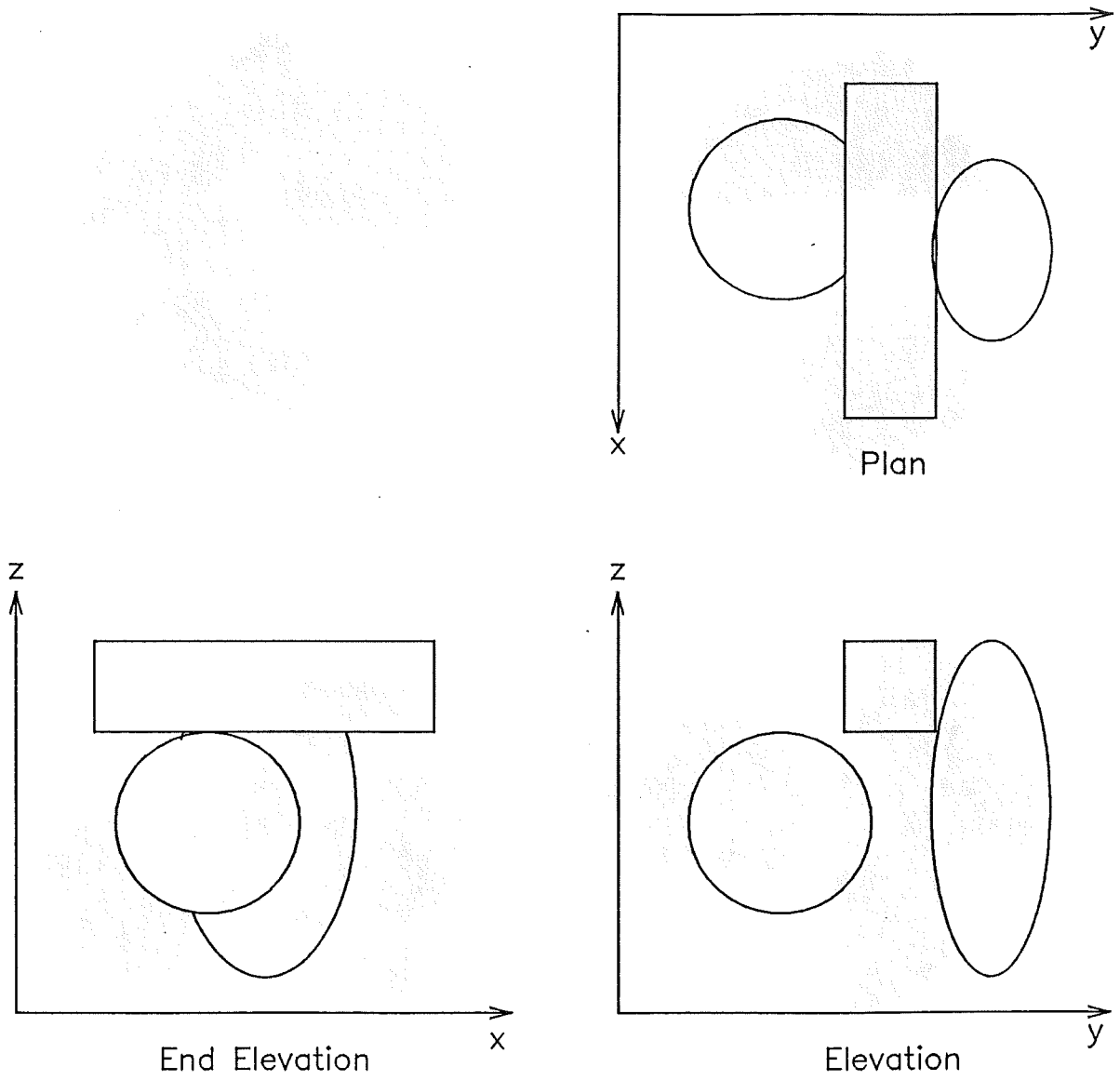


Figure 7.12: Orthographic drawing of the simulated object.

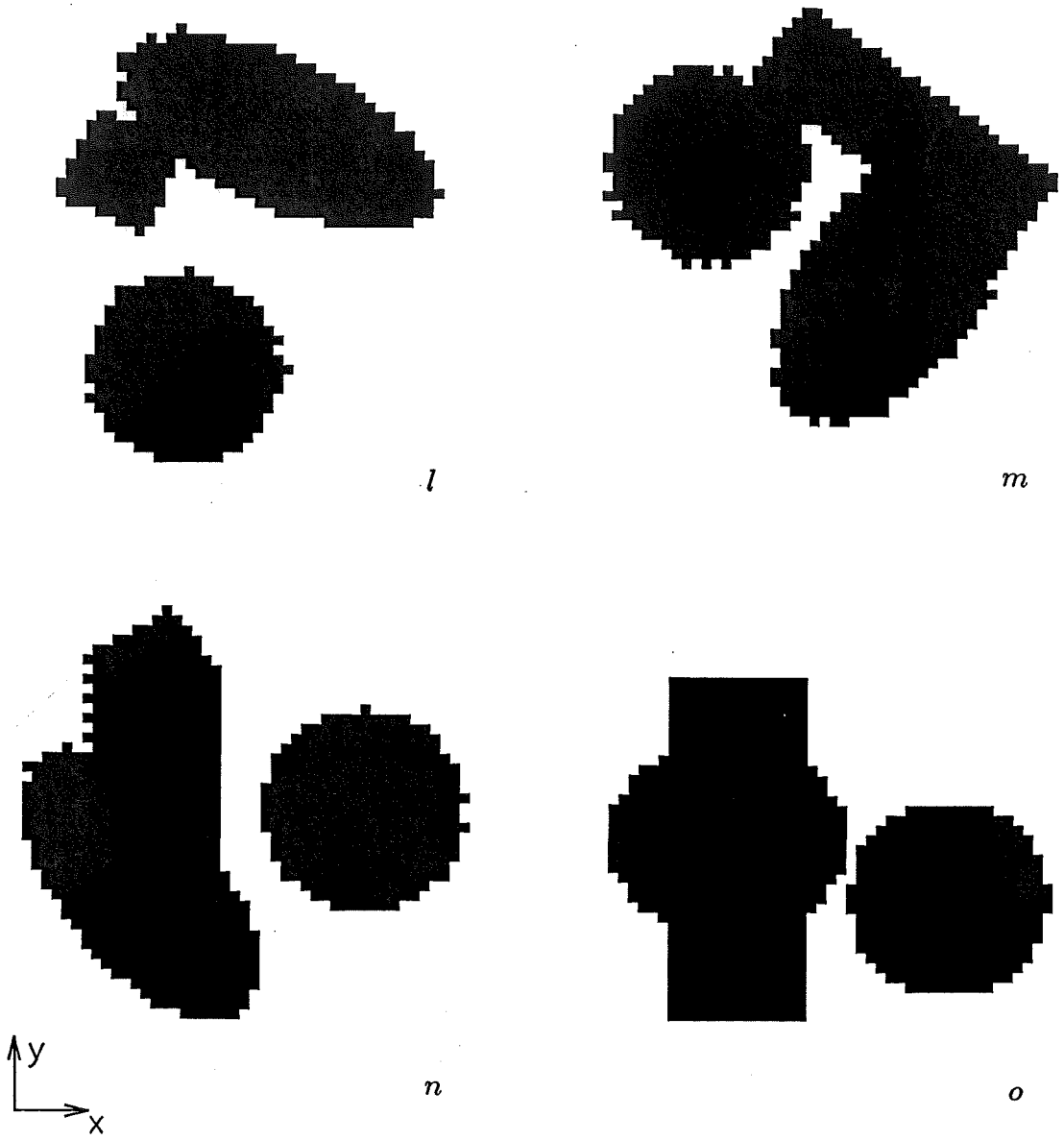


Figure 7.13: Four silhouettes of the simulated object shown in fig. 7.12. Respectively, the top left, top right, bottom left and bottom right silhouettes correspond to orientations of the object specified by $\alpha_l = -25°, \beta_l = -90°, \gamma_l = 0°$, $\alpha_m = -125°, \beta_m = -90°, \gamma_m = 46°$, $\alpha_n = 310°, \beta_n = -63°, \gamma_n = 159°$ and $\alpha_o = 0°, \beta_o = -35°, \gamma_o = 90°$.

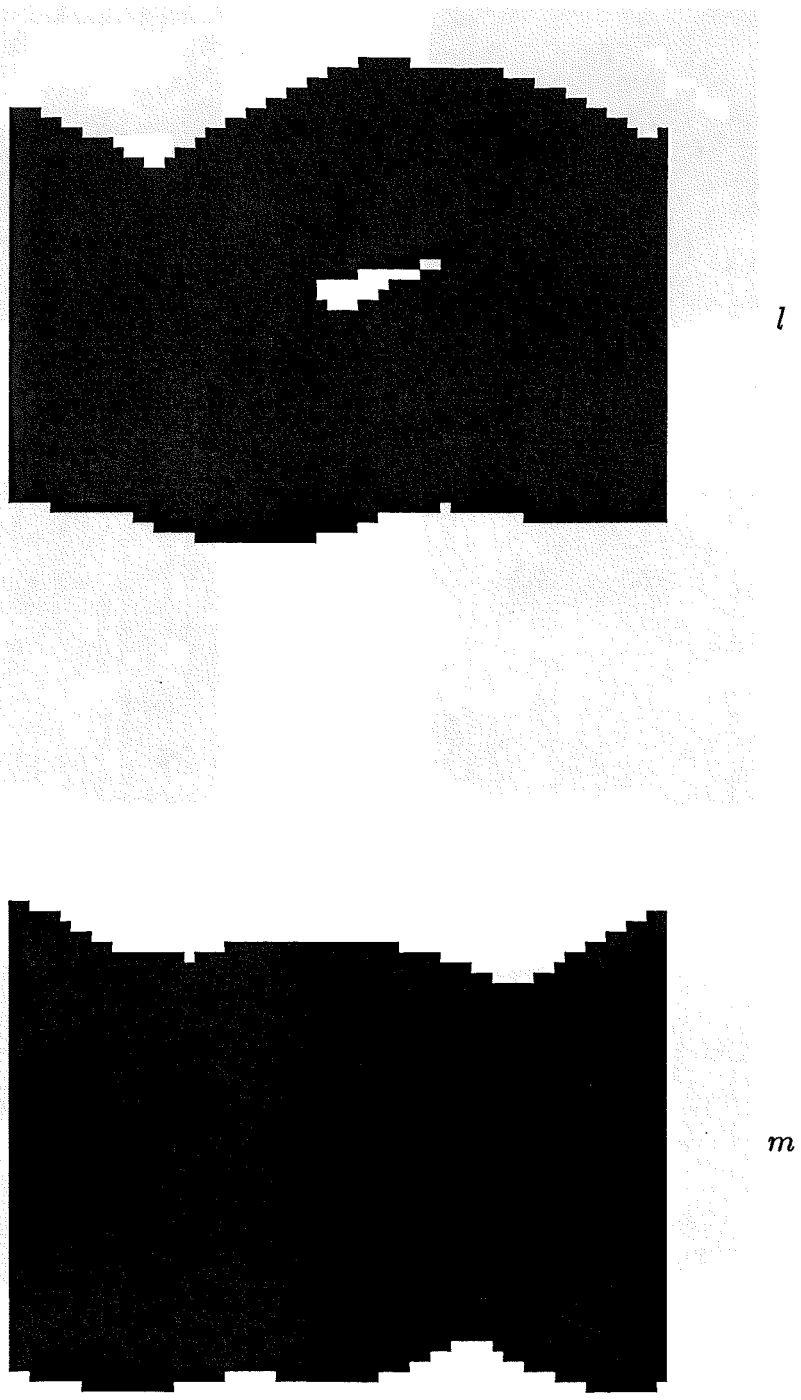


Figure 7.14: Binary sinograms of two of the silhouettes (those labelled l and m) shown in fig. 7.13.

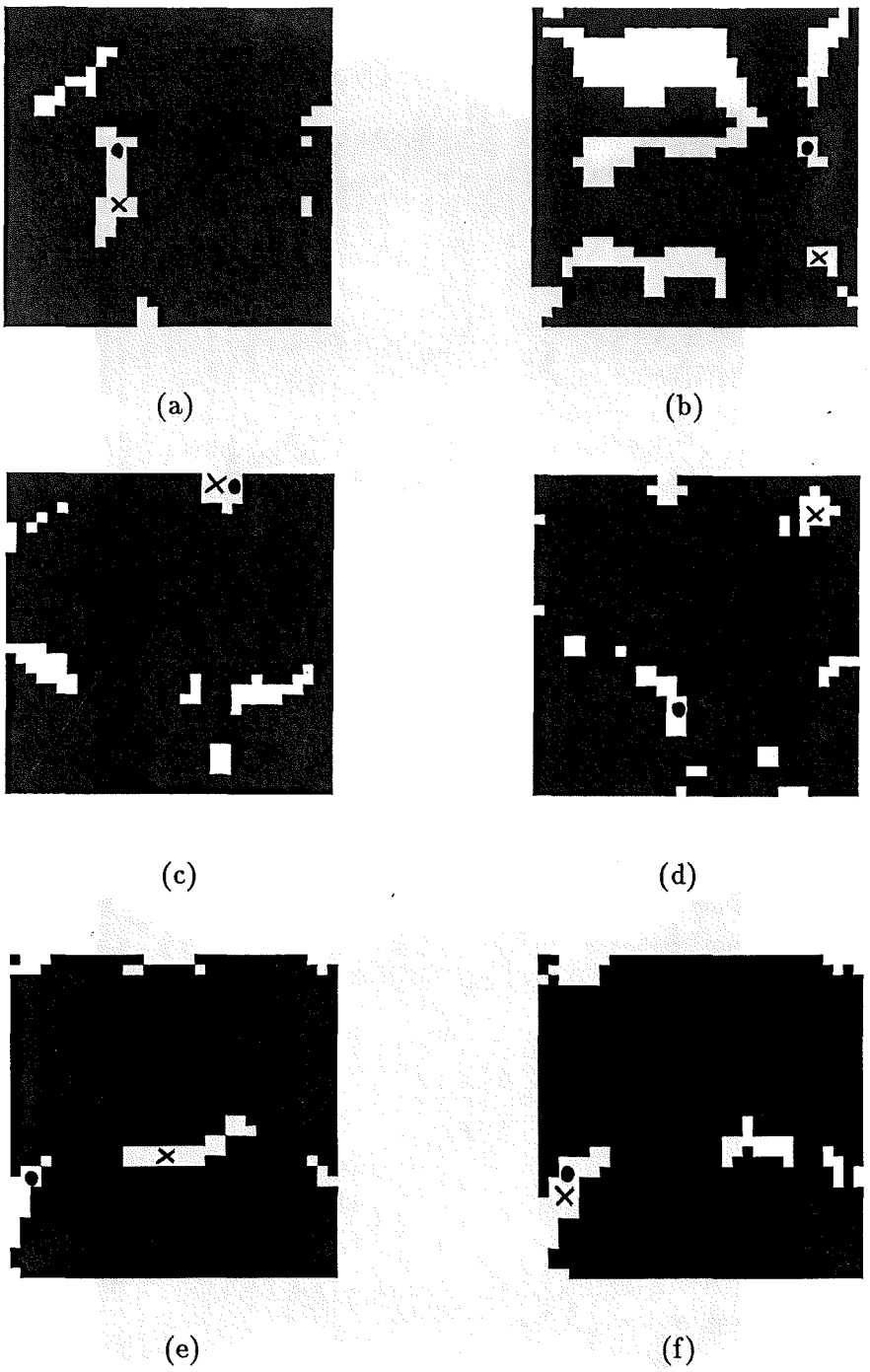


Figure 7.15: Images of the thresholded sccf functions (accepted points in white) resulting from the silhouettes in fig. 7.13. Each sub-figure corresponds to the invocation of (7.9) for the sinogram pairs; (a) l, m; (b) m, n; (c) n, l; (d) l, o; (e) m, o and (f) n, o. Crosses on the white regions identify the actual object orientations whilst the points marked • identify a trial orientation.

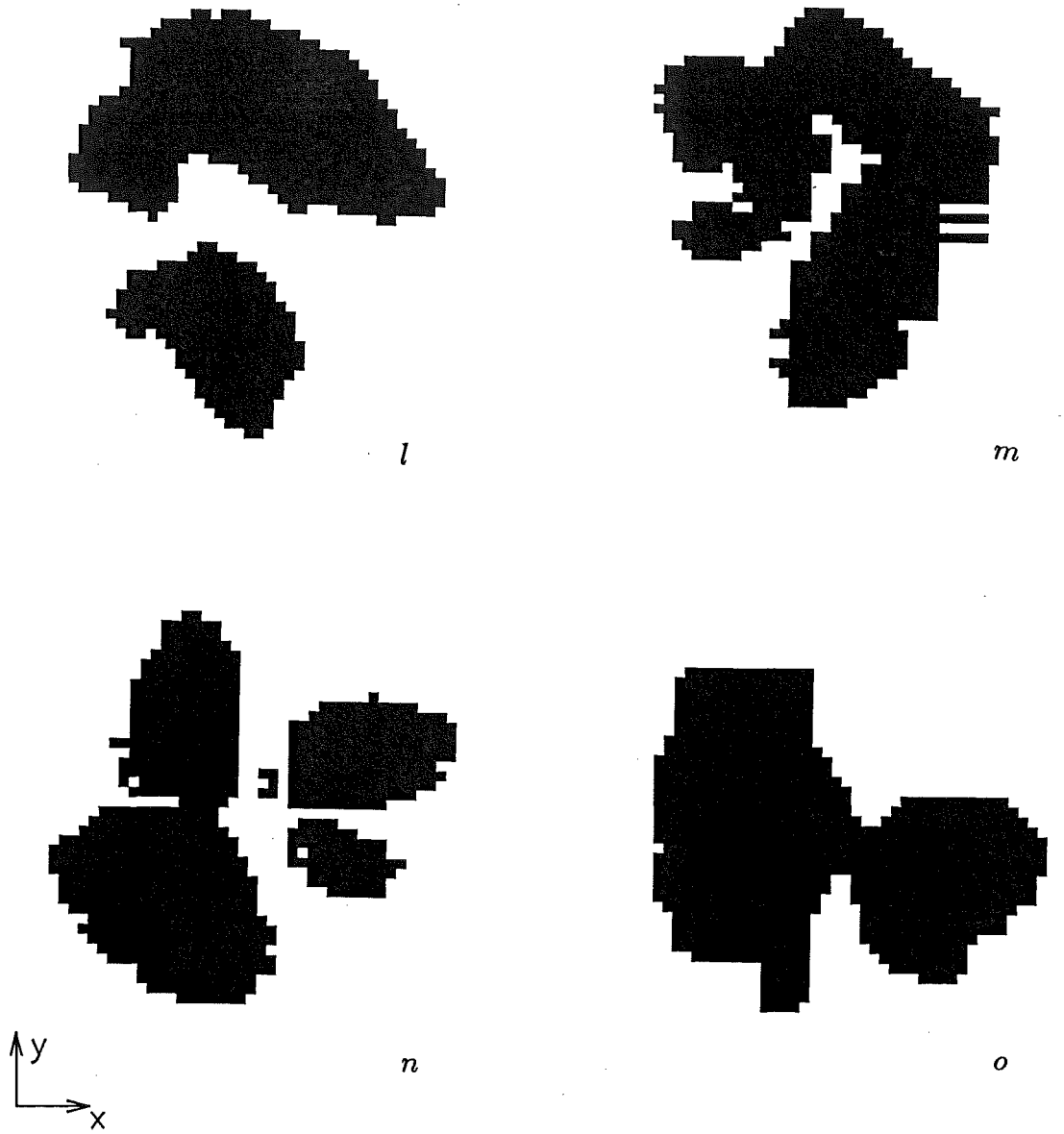
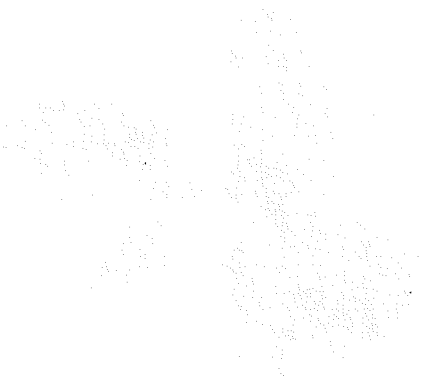


Figure 7.16: Reconstructed silhouettes of the simulated object (depicted in fig. 7.12) obtained by applying the self consistency check. The object is obtained by backprojecting the four silhouettes shown in fig. 7.13 at the orientations specified by the points marked • in fig. 7.15.



Chapter 8

Conclusions and suggestions for further research

The aim of this thesis has been to devise new or improved solutions to inverse problems, five of which have been investigated. These are: global solution to the inverse problem (§5.1), solution to an exterior inverse problem (§5.2), an extended Rayleigh-Gans (Born) approximation (§5.3), reconstruction of binary images from few projections (Chapter 6) and the determination of an unknown object's orientation and shape from silhouettes (Chapter 7).

Chapter 1 introduces the motivations for studying inverse problems. It further reviews established approaches to three types of inverse problem of considerable practical importance, in order to illuminate the present state of the theory behind inverse problems and to introduce common terminology and notation suitable for describing such problems. Chapter 1 also highlights the intimate relationship between direct and inverse problems. Equations describing wavefields and certain useful supplementary material are presented in Chapter 2. Solutions to direct scattering problems are reviewed in Chapter 3. Not only are some of these solutions invoked in later chapters, but they may also be suitable for incorporation into future approaches to solving inverse scattering problems. Chapter 4 reviews solutions to inverse scattering problems. Both exact and approximate solutions are examined.

This Chapter presents conclusions and suggestions for further research relating to the original results documented in Chapters 5 through 7. These conclusions and suggestions are collected into five sections, each relating to one of the abovementioned original contributions.

8.1 Global Solution to the Inverse Problem

§5.1 introduces an algorithm for reconstructing spatially variable refractive index distributions. The inversion algorithm is based on an expression (Bates, 1975) for the wave function which explicitly incorporates the inverse scattering data. It has been formulated for both two-dimensional and three-dimensional scatterers and has been implemented for two-dimensional scatterers.

While numerical results have only been obtained for scatterers having simple configurations, it is shown by example that convergence towards the correct Fourier coefficients of the refractive index distribution is achievable. It is also worth noting that the

sizes of the scatterers are significant, since a value for ka of unity and an average refractive index of 10.5 corresponds to strong scattering (as defined in §2.2.3). The algorithm is only one of two algorithms, the other being described by Colton and Monk (1988), based on exact formulations of scattering which have been demonstrated by example to exhibit convergence under strong scattering conditions. Furthermore, it is the only algorithm which has demonstrated convergence under strong scattering conditions when the refractive index distribution is asymmetrical.

Several aspects of the algorithm should be investigated to further improve its performance. Increasing the number of terms $\hat{\mu}$ in the Dini series allow $\psi(\mathbf{x}, k)$ to be represented more faithfully within Υ_-. Although this, in principle, allows one to increase the accuracy of the reconstructions, it tends to worsen the conditioning of (5.28). It may therefore prove worthwhile to attempt to devise strategies for optimising the value of $\hat{\mu}$. Calculating the Frèchet derivatives by central differences presently accounts for a significant fraction of the required computational effort. The numerical efficiency of the algorithm might be improved if some explicit means for calculating the Frèchet derivatives could be found.

A major current problem with numerical tests of any inverse scattering algorithm is the difficulty of generating accurate scattering data for objects of arbitrary shape and constitution. The scattering data invoked for the numerical examples presented in §5.1.4.2 could only be conveniently generated for specialised asymmetrical objects. What is now needed is a comprehensive attack on the direct problem, with the goal of devising an efficient, versatile and accurate algorithm. Further investigation of the accuracy of the algorithm introduced in §5.1 requires that scattering data be computed for a variety of asymmetric refractive index distributions.

It may also be fruitful to apply this algorithm to impenetrable and homogeneous objects to determine their shape $\sigma(\mathbf{x})$. (5.23) could then be reduced to a one-dimensional representation which should allow the algorithm to be formulated more simply and efficiently.

8.2 Solution to an Exterior Helmholtz Problem

In §5.2 the solution to the inverse problem for an impenetrable scatterer is formulated as a non-linear operator equation which is solved by the Newton-Kantorovich algorithm. The use of the null-field method to solve the direct problem leads to a manageable update procedure and quadratic convergence is demonstrated by example. Reconstructions generated by the algorithm appear to be more accurate than those reconstructed by other algorithms of this type (see §4.3.3).

Tables 5.6 and 5.7 suggest that the conditioning of the inverse problem, as defined by solution of the linear equations in the Newton-Kantorovich method, is such that the particular problem considered in §5.2 with the truncation values listed in §5.2.4, is not, in fact, ill-conditioned. The effect of noise on the data for this particular problem may not, therefore, be unduly significant.

It should be noted that this formulation and others (cf. Kristensson and Vogel 1986) that employ the null-field method to define Λ implicitly, assume that the origin of coordinates can be positioned at a point within Υ_-. Furthermore, since the effect of a shift of origin is to alter the phase of the scattered field whilst preserving its magnitude, the absolute position of the scatterer can only be determined when the phase of the

scattered field is known accurately. The problem of positioning the origin disappears when only the intensity of the scattered field is given. This is because an equivalence class of scatterers exists, consisting of translated versions of the true scatterer, whose members are all characterised by the same data. Consequently, when only the intensity of the scattered field is measured, only the shape of the scatterer can be deduced, and the experimenter is always free to select a representation in which the origin is within Υ_-.

The performance of the algorithm indicates that it is now sufficiently well developed to be applied to real-world scattering data. However, even though this algorithm is well-conditioned, the ever present contamination of the scattering data by noise and distortions is certain to require the incorporation of regularisation procedures into the algorithm.

8.3 An Extended Rayleigh-Gans (Born) Approximation

In §5.3 an approach to improving the performance of the Rayleigh-Gans (Born) approximation is introduced. The concrete manifestation of this approach is an algorithm which is as simple to implement as the conventional Rayleigh-Gans (Born) approximation. The numerical results presented in §5.3 suggest that this algorithm can in general be expected to generate explicit reconstructions which are significantly more accurate than those generated by the conventional Rayleigh-Gans (Born) approximation.

Several aspects of the new algorithm require further investigation. The algorithm needs to be applied to asymmetrical refractive index distributions. This would require the implementation of the back-propagation procedures introduced by Devaney (1985a) to avoid the introduction of interpolation artefacts into the reconstructions. An improved method of calculating the constant $\mathbf{a}_1$ introduced in (5.72) is also required. The inherent difficulty associated with calculating this constant is that the Fourier spectrum of $\chi^2(\mathbf{x}) - 1$ is only known on a circle whose origin and diameter are different for each k. It may be possible to overcome this difficulty by considering scattering data from a continuous range of k rather than just two values of k. It is also clear that, if means can be found for incorporating more than two terms into the distorting function, the performance of the algorithm should be enhanced. It may also prove possible to recover the extra terms from scattering data for a continuous range of k.

8.4 Reconstructing Binary Images from Few Projections

It is shown in Chapter 6 how to reconstruct a convex binary image from very few one-dimensional projections. In the absence of noise on the projections, accurate reconstructions can be generated from only two projections. It is worth remarking that binary images resulting from invocation of the modified back-projection algorithm (now the standard reconstruction procedure for x-ray computed tomography, see §4.2.1.1) from two projections are usually unrecognisable. The results also suggest that the algorithm can be expected to be usefully accurate when presented with noisy projections, with the accuracy improving as the number of given projections is increased. When rather more projections are given, but still far less than are needed to reconstruct faithful grey-scale images, it is shown in Chapter 6 that an extension to the algorithm can

reconstruct multiple-connected and non-convex binary images. Of course, the more intricate the shape of the image is, and the more noisy the projections, the more projections are needed to generate a reconstruction to whatever degree of faithfulness is required. Because the algorithm is free of multiplications, it is particularly efficient.

Although both forms of the algorithm introduced here can be readily implemented, and because computational experience seems to demonstrate that they are numerically stable, they would probably prove to be significantly more useful if they could be combined. It would seem, in particular, that the number of projections needed to reconstruct non-convex binary images might thereby be appreciably reduced. Additional significant improvement might result from employing several (say M) simultaneous starting points, in which case the value of $(K-1)$ would be M times that given by (6.8), suggesting that the effects of contamination might be further ameliorated.

It is noteworthy that the algorithms can function satisfactorily with arbitrarily oriented projections, provided of course that their angular separations are appreciable. This is yet another example demonstrating that missing projections can be compensated by *a priori* information, provided the latter is strong enough (Garden *et al.*, 1989). Knowing that the image is binary represents very strong *a priori* information of course.

8.5 Determining an Object's Orientation and Shape from Silhouettes

In Chapter 7 an algorithm is developed for reconstituting the orientation and shape of an object from three or more silhouettes. The relative orientations of the silhouettes need not be known. Neither is any *a priori* knowledge about the object required. The shape estimate can of course be expected to improve with increases in the number of given silhouettes of the object.

The algorithm could be adapted to estimating the amounts by which an object has been rotated between successive views. It could also be applied to determining the relative orientation of a known object from one silhouette having an unknown orientation.

A disadvantage of the algorithm is that it is highly computationally intensive. Even though the number of accepted points on the sccf functions is reduced by invocation of (7.8), many skeletons and resulting silhouettes must be submitted to the self consistency check. The number of times the self consistency check must be invoked is large because there are three sccf functions corresponding to each triple of silhouettes, in which every combination of accepted points on the sccf functions must be checked.

Approaches for reducing the computational effort required to implement the algorithm rest on increasing the speed of the hardware or making the algorithm more efficient. The most obvious means of making the algorithm more efficient is by somehow reducing the necessary number of accepted points on each sccf image. This could be achieved if the actual two-dimensional projections could be inferred from the silhouettes. Sinograms consisting of actual projections (as opposed to binary sinograms) could then be calculated, which would result in having far fewer accepted points on the sccf images. There are two possible ways of approximating the actual projections. The first is, essentially, to guess the actual projections from inferences drawn from the silhouettes. For instance, if a silhouette is narrow, it can reasonably be inferred that

the actual object is narrow, making it that much easier to estimate what the actual projection might be. Of course, the estimated projection may be so erroneous as to be virtually useless. The other, much more preferable way of estimating the projections, is to to supplement the silhouette of the object at any one orientation with several additional silhouettes at nearby orientations, with the relative orientations of these additional silhouettes being known. From these additional silhouettes more accurate inferences about the actual projections could then be made. Another advantage of this approach would be that more views of the object would be available for determining the object's shape. The algorithm's performance would be enhanced by implementing it in parallel processing hardware. The algorithm is particularly suited to parallel processing because each silhouette could be processed separately, implying that computational speed might be increased considerably by programming the algorithm for one of the recently developed computers possessing many parallel channels (Hwang, 1987; Miller and Stout, 1988; Cantoni and Levialdi, 1988).

References

ABRAMOWITZ, M. and STEGUN, I.A. (1965), *Handbook of Mathematical Functions*, Dover Publications, New York.

AGARWAL, G.S. (1976), 'Relation between Waterman's extended boudary condition and the generalised extinction theorem', *Physical Review D, Particles and Fields*, Vol. 14, No. 4, August, Pp. 1168–1171.

AGGARWAL, J.K. and MARTIN, W.N. (1983), 'Dynamic scene analysis', In HUANG, T. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 40–74.

AHUJA, N. and VEENSTRA, J. (1989), 'Generating octrees from object silhouettes in orthographic views', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-11, No. 2, February, Pp. 137–149.

ALVAREZ, R.E. and MACOVSKI, A. (1976), 'Energy-selective reconstructions in X-ray computerized tomography', *Physics in Medicine and Biology*, Vol. 21, No. 5, September, Pp. 733–744.

ANDERSEN, A.H. and KAK, A.C. (1982), 'Digital ray tracing in two-dimensional refractive fields', *Journal of the Acoustical Society of America*, Vol. 72, No. 5, November, Pp. 1593–1606.

ANGELL, T.S., KLEINMAN, R.E. and ROACH, G.F. (1987), 'An inverse transmission problem for the Helmholtz equation', *Inverse Problems*, Vol. 3, Pp. 149–180.

ARFKEN, G. (1985), *Mathematical Methods for Physicists*, Academic Press, New York, third ed.

ATKINSON, K.E. (1978), *An Introduction to Numerical Analysis*, Wiley, New York.

BAKER, B.B. and COPSON, E.T. (1950), *The Mathematical Theory of Huygens' Principle*, Clarendon, Oxford, second ed.

BALLARD, D.H. and BROWN, C.M. (1982), *Computer Vision*, Prentice-Hall, Englewood Cliffs,N.J.

BALTES, H.P. (Ed.) (1980), *Inverse Scattering Problems in Optics*, Springer-Verlag, Berlin. Topics in Current Physics volume 20.

BARNES, A. and SOLOMON, L.P. (1973), 'Some curious analytical ray paths for some interesting velocity profiles in geometrical acoustics', *Journal of the Acoustical Society of America*, Vol. 53, Pp. 147–155.

BARRETT, H.H. (1981), 'Three-dimensional image restoration from planar projections with application to optical data processing', In RHODES, W.T., FIENUP, J.R. and SALEH, B.E.A. (Eds.), *Proceedings of SPIE, 373, Transformations in Optical Signal Processing*, Pp. 179–190.

BATES, R.H.T. (1969), 'The theory of point matching for perfectly conducting waveguides and transmission lines', *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-17, Pp. 294–301.

BATES, R.H.T. (1975), 'Global solution to the scalar inverse scattering problem', *Journal of Physics A: Mathematical and General*, Vol. 8, No. 6, August, Pp. L80–82.

BATES, R.H.T. (1980), 'General introduction to the extended boundary condition', In VARADAN, V.K. and VARADAN, V.V. (Eds.), *Acoustic, Electromagnetic and Elastic scattering- focus on the T-matrix approach*, Pergamon Press, Pp. 21–31.

BATES, R.H.T. (1984), 'Full-wave computed tomography. Part 1: Fundamental theory', *IEE Proceedings Part A*, Vol. A131, No. 8, November, Pp. 610–615.

BATES, R.H.T. (1987), 'Aspects of the Erlangen-bispectrum', *Optik*, Vol. 76, No. 1, Pp. 23–26.

BATES, R.H.T. and MCDONNELL, M.J. (1989), *Image Restoration and Reconstruction*, Clarendon Press, Oxford, (corrected and updated reprinting).

BATES, R.H.T. and NG, F.L. (1972), 'Polarisation-source formulation of electromagnetism and dielectric-loaded waveguides', *Proceedings IEE*, Vol. 119, No. 11, November, Pp. 1568–1574.

BATES, R.H.T. and TAN, D.G.H. (1985), 'Towards reconstructing phases of inverse-scattering signals', *Journal of the Optical Society of America A*, Vol. 2, No. 11, November, Pp. 2013–2018. Special Issue on Inverse Problems in Propagation and Scattering.

BATES, R.H.T. and WALL, D.J.N. (1976), 'Chandrasekhar transformations improve convergence of computations of scattering from linearly stratified media', *IEEE Transactions on Antennas and Propagation*, Vol. AP-24, No. 2, March, Pp. 251–253.

BATES, R.H.T. and WALL, D.J.N. (1977), 'Null field approach to scalar diffraction: I. General method', *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, Vol. A287, No. 1339, September, Pp. 45–78.

BATES, R.H.T., BOERNER, W.M. and DUNLOP, R.G. (1976), 'An extended Rytov approximation and its significance for remote sensing and inverse scattering', *Optics Communications*, Vol. 18, No. 4, September, Pp. 421–423.

BATES, R.H.T., GARDEN, K.L. and PETERS, T.M. (1983), 'Overview of computerized tomography with emphasis on future developments', *Proceedings IEEE*, Vol. 71, No. 3, March, Pp. 356–372. Special Issue on Computerized Tomography.

BATES, R.H.T., MURCH, R.D. and SMITH, V.A. (1990), 'Manageable multidimensional inverse scattering theory', *Prepared by invitation for Physics Reports.*

BAYLISS, A. and TURKEL, E. (1980), 'Radiation boundary conditions for wave-like equations', *Communications on Pure and Applied Mathematics*, Vol. 33, Pp. 707–725.

BERG, O.R. and WOOLVERTON, D.G. (Eds.) (1986), *Seismic Stratigraphy II: An integrated Approach to Hydrocarbon Exploration*, The American Association of Petroleum Geologists, Tulsa, Oklahoma, U.S.A., 2 ed.

BERNTSEN, S. (1983), 'Diffraction of an electric polarized wave by a dielectric wedge', *SIAM Journal of Applied Mathematics*, Vol. 43, No. 1, February, Pp. 186–211.

BERTERO, M. and MOL, C.D. (1981), 'Stability problems in inverse diffraction', *IEEE Transactions on Antennas and Propagation*, Vol. AP-29, No. 2, March, Pp. 368–372.

BERTERO, M., BRIANZI, P., PIKE, E.R. and REBOLIA, L. (1988), 'Linear regularizing algorithms for positive solutions of linear inverse problems', *Proceedings of the Royal Society of London A: Mathematical and Physical Sciences*, Vol. 415, Pp. 257–275.

BEUSMANS, J.M.H., HOFFMAN, D.D. and BENNET, B.M. (1987), 'Description of solid shape and its inference from occluding contours', *Journal of the Optical Society of America A*, Vol. 4, No. 7, July, Pp. 1155–1167.

BEYLKIN, G. (1985a), 'Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform', *Journal of Mathematical Physics*, Vol. 26, No. 1, January, Pp. 99–108.

BEYLKIN, G. (1985b), 'Reconstructing discontinuities in multidimensional inverse scattering problems: Smooth errors vs small errors', *Applied Optics*, Vol. 24, No. 23, dec, Pp. 4086–4088.

BILLINGSLEY, J. (Ed.) (1985), *Robots and Automated Manufacture*, Peter Peregrins Ltd for I.E.E., London.

BLACKLEDGE, J.M., BURGE, R.E., HOPCRAFT, K.I. and WOMBELL, R.J. (1987), 'Quantitative diffraction tomography: I. Pulsed acoustic fields', *Journal of Physics D : Applied Physics*, Vol. 20, No. 1, January, Pp. 1–10.

BLAKE, L.V. (1984), *Antennas*, Artech House, Inc., Norwood, MA, USA, 2nd ed. Originally published by John Wiley & Sons, 1966.

BLOHBAUM, J. (1989), 'Optimisation methods for an inverse problem with time-harmonic electromagnetic waves: An inverse problem in electromagnetic scattering', *Inverse Problems*, Vol. 5, Pp. 463–482.

BORN, M. and WOLF, E. (1970), *Principles of Optics*, Pergamon Press, fourth ed.

BOWMAN, J.J., SENIOR, T.B.A. and USLENGHI, P.L.E. (Eds.) (1969), *Electromagnetic and Acoustic Scattering by Simple Shapes*, North-Holland, Amsterdam.

BOYSE, W.E. and KELLER, J.B. (1986), 'Inverse elastic scattering in three dimensions', *Journal of the Acoustical Society of America*, Vol. 79, No. 2, February, Pp. 215–218.

BRACEWELL, R.N. (1978), *The Fourier Transform and its Applications*, McGraw-Hill, New York, second ed.

BRADY, M. (1982), 'Computational approaches to image understanding', *ACM Computing Surveys*, Vol. 14, No. 1, March, Pp. 3–71.

BRADY, M. and YUILLE, A. (1984), 'An extremum principle for shape from contour', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 3, May, Pp. 288–301.

BRAGGINS, D. and HOLLINGUM, J. (1986), *The Machine Vision Sourcebook*, Springer-Verlag.

BRIGHAM, E.O. (1974), *The Fast Fourier Transform*, Prentice-Hall, Englewood Cliffs, New Jersey.

BRUCKSTEIN, A.M. and KAILATH, T. (1987), 'Inverse scattering for discrete transmission-line models', *SIAM Review*, Vol. 29, No. 3, September, Pp. 359–389.

BUSINGER, P.A. and GOLUB, G.H. (1969), 'Algorithm 358: Singular value decomposition of a complex matrix', *Communications of the Association for Computing Machinery*, Vol. 12, No. 10, October, Pp. 564–565.

CANTONI, V. and LEVIALDI, S. (1988), 'Multiprocessor computing for images', *Proceedings IEEE*, Vol. 76, No. 8, August, Pp. 959–981.

CHA, S. and VEST, C.M. (1981), 'Tomographic reconstruction of strongly refracting fields and its application to interferometric measurement of boundary layers', *Applied Optics*, Vol. 20, No. 16, aug, Pp. 2787–2794.

CHADAN, K. and SABATIER, P.C. (1977), *Inverse Problems in Quantum Scattering Theory*, Texts and Monographs in Physics, Springer-Verlag, New York.

CHERIOT, T.E. (1971), 'Calculation of near field of circular aperture antenna using geometrical theory of diffraction', *IEEE Transactions on Electromagnetic Compatability*, Vol. EMC-13, No. 2, May, Pp. 29–34.

CHERNOV, L. (1967), *Wave Propagation in a Random Medium*, Dover, New York.

CHIN, R.T. and DYER, C.R. (1986), 'Model-based recognition in robot vision', *ACM Computing Surveys*, Vol. 18, No. 1, March, Pp. 67–108.

CLAYTON, R.W. and STOLT, R.H. (1981), 'A Born-WKBJ inversion method for acoustic reflection data', *Geophysics*, Vol. 46, No. 11, November, Pp. 1559–1567.

COLTON, D. (1984), 'The inverse scattering problem for time-harmonic acoustic waves', *SIAM Review*, Vol. 26, No. 3, July, Pp. 323–350.

COLTON, D. and KRESS, R. (1983), *Integral Equation Methods in scattering theory*, John Wiley, New York.

COLTON, D. and MONK, P. (1985), 'A novel method for solving the inverse scattering problem for time-harmonic acoustic waves in the resonance region', *SIAM Journal of Applied Mathematics*, Vol. 45, No. 6, December, Pp. 1039–1053.

COLTON, D. and MONK, P. (1986), 'A novel method for solving the inverse scattering problem for time-harmonic acoustic waves in the resonance region II', *SIAM Journal of Applied Mathematics*, Vol. 46, No. 3, December, Pp. 506–523.

COLTON, D. and MONK, P. (1987), 'The inverse scattering problem for time-harmonic acoustic waves in a penetrable medium', *Quarterly Journal of Mechanics and Applied Mathematics*, Vol. 40, No. 2, May, Pp. 184–212.

COLTON, D. and MONK, P. (1988), 'The inverse scattering problem for time-harmonic acoustic waves in an inhomogeneous medium', *Quarterly Journal of Mechanics and Applied Mathematics*, Vol. 41, No. 1, Pp. 97–125.

CONNOLLY, T.J. (1989), *Nonlinear Methods for inverse problems*, PhD thesis, Department of Mathematics, University of Canterbury, Christchurch, New Zealand.

CONNOLLY, C.I., MUNDY, J.L., STENSTROM, J.R. and THOMPSON, D.W. (1987), 'Matching from 3-D models into 2-D intensity scenes', In *Proceedings of First International Conference on Computer Vision*, Computer Society Press of the IEEE, Pp. 65–73.

CONTE, S.D. and DE BOOR, C. (1980), *Elementary Numerical Analysis - An algorithmic approach*, Wiley and sons, New York, third ed.

COOLEY, J.W. and TUKEY, J.W. (1965), 'An algorithm for the machine calculation of complex Fourier variables', *Mathematics of Computation*, Vol. 19, No. 297.

CRAIG, I.J.D. and BROWN, J.C. (1986), *Inverse problems in astronomy*, Adam Hilger, Bristol and Boston.

DAILY, W.D. (Ed.) (1986), 'Special section on geotomography', *Proceedings IEEE*, Vol. 74, No. 2, February, Pp. 243–361.

DENNIS, J.E. and SCHNABEL, R.B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall Series in Computational Mathematics, Prentice-Hall, Engelwood Cliffs, New Jersey.

DEVANEY, A.J. (1982), 'A filtered backpropagation algorithm for diffraction tomography', *Ultrasonic Imaging*, Vol. 4, No. 4, October, Pp. 336–350.

DEVANEY, A.J. (1983), 'A computer simulation study of diffraction tomography', *IEEE Transactions on Biomedical Engineering*, Vol. BME-30, No. 7, July, Pp. 377–386. Special Issue on Medical Ultrasound.

DEVANEY, A.J. (1984), 'Geophysical diffraction tomography', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-22, No. 1, January, Pp. 3–13.

DEVANEY, A.J. (1985a), 'Diffraction tomography', In BOERNER, W.M. (Ed.), *Inverse Methods in Electromagnetic Imaging*, D. Reidel Publishing Company, Pp. 1107–1135.

DEVANEY, A.J. (1985b), 'Generalized projection-slice theorem for fan beam diffraction tomography', *Ultrasonic Imaging*, Vol. 7, No. 3, July, Pp. 264–275.

DEVANEY, A.J. (1986), 'Reconstructive tomography with diffracting wavefields', *Inverse Problems*, Vol. 2, No. 2, May, Pp. 161–183.

DEVANEY, A.J. (1987), 'A fast filtered backpropagation algorithm for ultrasound tomography', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-34, No. 3, May, Pp. 330–339.

DEVANEY, A.J. (1989), 'The limited-view problem in diffraction tomography', *Inverse Problems*, Vol. 5, Pp. 501–521.

DEVANEY, A.J. and ORISTAGLIO, M.L. (1983), 'Inversion procedure for inverse scattering within the distorted-wave Born approximation', *Physical review letters*, Vol. 51, No. 4, July, Pp. 237–240.

DEVANEY, A.J. and SHERMAN, G.C. (1982), 'Non-uniqueness in inverse source and scattering problems', *IEEE Transactions on Antennas and Propagation*, Vol. AP-30, No. 5, September, Pp. 1034–1037.

DINES, K.A. and GOSS, S.A. (1987), 'Computed ultrasonic reflection tomography', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-34, No. 3, may, Pp. 309–318.

DUDA, R.O. and HART, P.E. (1973), *Pattern classification and scene analysis*, John Wiley and Sons, New York.

DUNLOP, G.R. (1978), *Ultrasonic Transmission Imaging*, PhD thesis, University of Canterbury, New Zealand.

DUNLOP, G.R., BOERNER, W.M. and BATES, R.H.T. (1976), 'On an extended Rytov approximation and its comparison with the Born approximation', In *IEEE AP-S International Symposium*, Pp. 587–591.

FATEMI, C.N.L.M. and WAAG, R.C. (1983), 'Digital processing for improvement of ultrasonic abdominal images', *IEEE Transactions on Medical Imaging*, Vol. MI-2, No. 2, June, Pp. 67–75.

FAWCETT, J. and KELLER, H.B. (1985), 'Three-dimensional ray tracing and geophysical inversion in layered media', *SIAM Journal of Applied Mathematics*, Vol. 45, No. 3, June, Pp. 492–501.

FELSON, L.B. and MARCUVITZ, N. (1973), *Fields and Waves*, Penguium, New York, 1 ed.

FLETCHER, R. (1980), *Practical Methods of Optimization*, Vol. 1: Unconstrained Optimization, Wiley, Chichester, England.

FRANKLIN, P. (1960), in *Fundamental Formulas of Physics*, (D. H. Menzel ed) Dover, N.Y.

FU, K.S. (1982), *Syntactic Pattern Recognition and Applications*, Prentice Hall, Englewood Cliffs, New Jersey.

FUKUSHIMA, K. (1988), 'A neural network for visual pattern recognition', *IEEE Computer*, Vol. 21, No. 3, March, Pp. 65–75.

GARABEDIAN, P.R. (1964), *Partial Differential Equations*, Wiley, New York.

GARDEN, K.L. (1984), *An Overview of Computed Tomography*, PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand.

GARDEN, K.L. and BATES, R.H.T. (1984), 'Image reconstruction from projections VII: interactive reconstruction of piecewise constant images from few projections', *Optik*, Vol. 68, No. 2, September, Pp. 161–173.

GARDEN, K.L., BATES, R.H.T., WON, M.C. and CHIKWANDA, H. (1984), 'Computerized tomographic imaging is insensitive to density variation during scanning', *Image and Vision Computing*, Vol. 2, No. 2, May, Pp. 76–84.

GARDEN, K.L., BONES, P.J. and BATES, R.H.T. (1989), 'From living being to medical image - bridging the dimensionality gap', *Australasian Physical and Engineering Sciences in Medicine*, Vol. 12, No. 4, December, Pp. 186–204.

GARDENIER, P.H., MCCALLUM, B.C. and BATES, R.H.T. (1986), 'Fourier transform magnitudes are unique to pattern recognition templates', *Biological Cybernetics*, Vol. 54, Pp. 385–391.

GELFAND, I.M. and LEVIATAN, B.M. (1955), 'On the determination of a differential equation from its spectral function', *Soviet Mathematics (American Mathematical Society translation of Doklady, Academy of Sciences of the USSR)*, Vol. 1, Pp. 253–304.

GILBERT, P. (1972), 'Iterative methods for the three-dimensional reconstruction of an object from projections', *Journal of Theoretical Biology*, Vol. 36, No. 1, July, Pp. 105–117.

GLOVER, G.H. and SHARP, J.C. (1977), 'Reconstruction of ultrasound propagation speed distributions in soft tissue: Time-of-flight tomography', *IEEE Transactions on Sonics and Ultrasonics*, Vol. SU-24, No. 4, July, Pp. 229–234.

GOLUB, G.H. and VAN LOAN, C.F. (1983), *Matrix Computations*, North Oxford Academic, Oxford.

GOODMAN, J.W. (1968), *Introduction to Fourier Optics*, McGraw-Hill, San Francisco.

GOODMAN, J.W. (1975), 'Statistical properties of laser speckle patterns', In DAINTY, J.C. (Ed.), *Laser Speckle and Related Phenomena*, Springer-Verlag, Berlin, Chap. 2, Pp. 9–75.

GREENLEAF, J.F. (1983), 'Computerized tomography with ultrasound', *Proceedings IEEE*, Vol. 71, No. 3, March, Pp. 330–337. Special Issue on Computerized Tomography.

GRIMSON, W.E.L. (1987), 'Recognition of object families using parameterised models', In *Proceedings of First International Conference on Computer Vision*, Computer Society Press of the IEEE, Pp. 93–101.

HALLIWELL, M. (1987), 'Ultrasonic imaging medical diagnosis', *IEE Proceedings Part A*, Vol. 134, No. 2, Febrary, Pp. 129–135.

HARRINGTON, R.F. (1968), *Field Computation by Moment Methods*, MacMillan, New York.

HASHABY, T.M. and MITTRA, R. (1987), 'Review of some inverse methods in electromagnetics', *Journal of the Optical Society of America A*, Vol. 4, No. 1, January, Pp. 281–291. Special Issue on Signal Recovery.

HEADING, J. (1962), *An Introduction to Phase-integral Methods*, John Wiley and sons, New York.

HERMAN, G.T. (Ed.) (1979), *Image Reconstruction from Projections. Implementation and Applications*, Topics in Applied Physics volume 32, Springer-Verlag, Berlin.

HERMAN, G.T. (1980), *Image Reconstruction from Projections*, Computer Science and Applied Mathematics, Academic Press, New York.

HERMAN, G.T. (Ed.) (1983), 'Special issue on computerized tomography', *Proceedings IEEE*, Vol. 71, No. 3, March, Pp. 291–435.

HERRMANN, G.F. (1987), 'Numerical computation of diffraction coefficients', *IEEE Transactions on Antennas and Propagation*, Vol. AP-35, No. 1, January, Pp. 53–61.

HORN, B.K.P. (1975), 'Obtaining shape from shading information', *In Psychology of Computer Vision, (P. H. Winston ed.)*, Pp. 115–155.

HUANG, T.S. (Ed.) (1981), *Image sequence analysis*, Springer-Verlag, Berlin. Springer series in information sciences.

HUNGWEN, L. and KENDER, J.R. (Eds.) (1988), 'Special issue on computer vision', *Proceedings IEEE*, Vol. 76, No. 8, August, Pp. 863–1050.

HWANG, K. (1987), 'Advanced parallel processing with supercomputer architectures', *Proceedings IEEE*, Vol. 75, No. 10, October, Pp. 1348–1379.

IKEUCHI, K. and HORN, B.K.P. (1981), 'Numerical shape from shading and occluding boundaries', *Artificial Intelligence*, Vol. 17, August, Pp. 141–184.

JAMES, B.D. (1986), *Vector Diffraction Tomography Using Radon Transform Techniques in Computer Assisted Electromagnetic Imaging*, Master's thesis, Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, July.

JOHNSON, S.A. and TRACY, M.L. (1983), 'Inverse scattering solutions by a sinc basis, multiple source, moment method — Part I: Theory', *Ultrasonic Imaging*, Vol. 5, No. 4, October, Pp. 361–375.

JOHNSON, S.A., ZHOU, Y., TRACY, M.K., BERGGREN, M.J. and STENGER, F. (1984), 'Inverse scattering solutions by a sinc basis, multiple source, moment method — Part III: Fast algorithms', *Ultrasonic Imaging*, Vol. 6, No. 1, January, Pp. 103–116.

JONES, D.S. (1964), *The Theory of Electromagnetism*, Pergamon Press, New York.

JONES, D.S. (1974), 'Integral equations for the exterior acoustic problem', *Quarterly Journal of Mechanics and Applied Mathematics*, Vol. 27, Pp. 129–142.

JONES, D.S. (1979), *Methods in Electromagnetic Wave Propagation*, Oxford University Press, Oxford.

JONES, D.S. (1988a), 'Surface radiation conditions', *IMA Journal of applied mathematics*, Vol. 41, Pp. 21–30.

JONES, D.S. (1988b), 'An approximate boundary condition in acoustics', *Journal of sound and vibration*, Vol. 121, No. 1, Pp. 37–45.

JONES, D.S. and MAO, X.Q. (1989), 'The inverse problem in hard acoustic scattering', *Inverse Problems*, Vol. 5, Pp. 731–748.

KAK, A.C., VAYDA, A.J., CROMWELL, R.L., KIM, W.Y. and CHEN, C.H. (1988), 'Knowledge-based robotics', *International Journal of Production Research*, Vol. 26, No. 5, May, Pp. 707–734.

KELLER, J.B. (1958), 'The inverse scattering problem in geometrical optics and the design of reflectors', *IRE Transactions on Antennas and Propagation*, April, Pp. 146–149.

KELLER, J.B. (1962), 'Geometrical theory of diffraction', *Journal of the Optical Society of America*, Vol. 52, No. 2, Pp. 116–130.

KELLER, J.B. (1969), 'Accuracy and validity of the Born and Rytov approximations', *Journal of the Optical Society of America*, Vol. 59, No. 8, August, Pp. 1003–1004.

KEMPER, A. and WALLRATH, M. (1987), 'An analysis of geometric modelling in database systems', *ACM computing Surveys*, Vol. 19, No. 1, March, Pp. 47–91.

KIM, J.H., SONG, T.K. and PARK, S.B. (1987), 'Pipelined sampled-delay focusing in ultrasound imaging systems', *Ultrasonic Imaging*, No. 9, Pp. 75–91.

KINBER, B.Y. (1962), 'On diffraction by the open end of a sectorial horn', *Radioteckh. Elektron*, Vol. 7, Pp. 1749–1762.

KIRSCH, A. and KRESS, R. (1987), *Inverse Problems*, Academic, New York, Pp. 274–290.

KIRSCH, A., KRESS, R., MONK, P. and ZINN, A. (1988), 'Two methods for solving the inverse acoustic scattering problem', *Inverse Problems*, Vol. 4, Pp. 749–770.

KLEINMAN, R.E. and ROACH, G.F. (1974), 'Boundary integral equations for the three-dimensional Helmholtz equation', *SIAM Review*, Vol. 16, Pp. 214–236.

KOUYOUMJIAN, R.G. (1965), 'Asymptotic high-frequency methods', *Proceedings IEEE*, Vol. 53, August, Pp. 864–876.

KOUYOUMJIAN, R.G. and PATHAK, P.H. (1974), 'A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface', *Proceedings IEEE*, Vol. 62, No. 11, November, Pp. 1448–1461.

KRIEGSMANN, G.A. and MORAWETZ, C. (1980), 'Solving the Helmholtz equation for exterior problems with variable refractive index of refraction', *SIAM Journal on Scientific and Statistical Computing*, September, Pp. 371–385.

KRIEGSMANN, G.A., TAFLOVE, A. and UMASHANKAR, K.R. (1987), 'A new formulation of electromagnetic wave scattering using an on-surface radiation boundary condition approach', *IEEE Transactions on Antennas and Propagation*, Vol. AP-35, No. 2, February, Pp. 153–161.

KRISTENSSON, G. and VOGEL, C.R. (1986), 'Inverse problems for acoustic waves using the penalised likelihood method', *Inverse Problems*, Vol. 2, No. 4, November, Pp. 461–479.

LEE, S. and DESCHAMPS, G.A. (1976), 'A uniform asymptotic theory of electromagnetic diffraction by a curved wedge', *IEEE Transactions on Antennas and Propagation*, Vol. AP-24, No. 1, January, Pp. 25–34.

LEEMAN, S., FIDDY, M.A. and ZAPALOWSKI, L. (1985), 'Born versus Rytov: is the debate over?', In DEVANEY, A.J. and BATES, R.H.T. (Eds.), *Inverse Optics II*, August, Pp. 11–14. Proceedings SPIE volume 558 (the Society of Photo-Optical Instrumentation Engineers).

LESSELIER, D., VUILLET-LAURENT, D., JOUVIE, F. and TABBARA, W. (1985), 'Iterative solution of some direct and inverse problems in electromagnetics and acoustics', *Electromagnetics*, Vol. 5, No. 2-3, Pp. 147–189.

LEVY, B.C. and ESMERSOY, C. (1988), 'Variable background Born inversion by wavefield backpropagation', *SIAM Journal of Applied Mathematics*, Vol. 48, No. 5, October.

LEWIN, L. (1970), 'On the restricted validity of point-matching techniques', *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-18, No. 12, December, Pp. 1041–1047.

LEWITT, R.M. (1983), 'Reconstruction algorithms: Transform methods', *Proceedings IEEE*, Vol. 71, No. 3, March, Pp. 390–408.

LEWITT, R.M. and BATES, R.H.T. (1978), 'Image reconstruction from projections I: General theoretical considerations', *Optik*, Vol. 50, No. 1, February, Pp. 19–33.

LINES, L.R. (Ed.) (1986), 'Special issue on seismic inversion', *Proceedings IEEE*, Vol. 74, No. 3, March, Pp. 387–509.

LO, T., TOKSÖZ, M.N., XU, S. and WU, R. (1988), 'Ultrasonic laboratory tests of geophysical tomographic reconstruction', *Geophysics*, Vol. 53, No. 7, July, Pp. 947–955.

LOWE, D.G. (1987), 'Three-dimensional object recognition from single two-dimensional images', *Artificial Intelligence*, Vol. 31, No. 3, March, Pp. 355–395.

LYTLE, R.J. and DINES, K.A. (1980), 'Iterative ray-tracing between boreholes for underground image reconstruction', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-18, No. 3, July, Pp. 234–240.

MADSEN, K. and SCHJAER-JACOBSEN, H. (1978), 'Linearly constrained minimax optimisation', *Mathematical Programming*, Vol. 14, Pp. 208–223.

MCGREGOR, B.P. (1988), *Angular Reconstitution of Silhouettes*, Third Professional Year Project report, Electrical Engineering Department, University of Canterbury, New Zealand.

MCKINNON, G.C. and BATES, R.H.T. (1980), 'A limitation on ultrasonic transmission tomography', *Ultrasonic Imaging*, Vol. 2, No. 1, January, Pp. 48–54.

MCQUILLIN, R., BACON, M. and BARCLAY, W. (1984), *An Introduction to seismic interpretation*, Graham and Trotman, Oxford, Great Britain.

MILLER, R. and STOUT, Q.F. (1988), 'Efficient parallel convex hull algorithms', *IEEE Transactions on Computers*, Vol. CO-37, No. 12, December, Pp. 1605–1618.

MITTRA, R. (Ed.) (1973), *Computer Techniques for Electromagnetics*, Vol. 7 of Electrical Engineering, Pergamon Press, New York.

MITTRA, R., KO, W.L. and RAHMAT-SAMII, Y. (1979), 'Transform approach to electromagnetic scattering', *Proceedings IEEE*, Vol. 67, No. 11, November, Pp. 1486–1503.

MORSE, P.M. and FESHBACH, H. (1953), *Methods of Theoretical Physics*, McGraw-Hill, New York. Volume 1 : Chapters 1 to 8; volume 2 : Chapters 9 to 13.

MOSS, A. (1982), 'Imaging modalities', *IEEE Transactions on Medical Imaging*, Vol. MI-1, No. 4, December, Pp. 215–219.

MUELLER, R.K., KAVEH, M. and WADE, G. (1979), 'Reconstructive tomography and applications to ultrasonics', *Proceedings IEEE*, Vol. 67, Pp. 567–587.

MURCH, R.D. and BATES, R.H.T. (1990a), 'Image reconstruction from projections IX: Binary images'. submitted to Optik.

MURCH, R.D. and BATES, R.H.T. (1990b), 'Image reconstruction from projections XI: reconstituting object shape and orientation'. submitted to Optik.

MURCH, R.D., TAN, D.G.H. and WALL, D.J.N. (1988a), 'Newton-Kantorovich method applied to the two-dimensional inverse scattering for an exterior Helmholtz problem', *Inverse Problems*, Vol. 4, No. 4, Pp. 1117–1128.

MURCH, R.D., QUEK, B.K. and MCGREGOR, B.M. (1988b), 'Determination of an objects orientation from silhouettes', In *Proceedings of the New Zealand National Electronics Conference*, Christchurch, 31 August - 2 September, Pp. 59–64.

NAGEL, H.H. (1983), 'Overview on image sequence analysis', In HUANG, T. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 2–39.

NASHED, M.Z. (1981), 'Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory', *IEEE Transactions on Antennas and Propagation*, Vol. AP-29, No. 2, March, Pp. 220–231. Special Issue on Inverse Methods in Electromagnetics.

NEWTON, R.G. (1981), 'Inversion of reflection data for layered media: A review of exact methods', *Geophysical Journal of the Royal Astronomical Society*, Vol. 65, No. 1, April, Pp. 191–215.

NORTON, S.J. (1987), 'Computing ray trajectories between two points: A solution to the ray-linking problem.', *Journal of the Optical Society of America A*, Vol. 4, No. 10, October, Pp. 1919–1922.

NUDELMAN, S. and PATTON, D. (Eds.) (1980), *Imaging for Medicine*, Vol. 1, Plenum Press, New York.

OCHS, R. (1987), 'The limited aperture problem of inverse acoustic scattering: Dirichlet boundary conditions', *SIAM Journal of Applied Mathematics*, Vol. 47, No. 6, December, Pp. 1320–1341.

O'DONNELL, M. (1988), 'Applications of VLSI circuits to medical imaging', *Proceedings IEEE*, Vol. 76, No. 9, September, Pp. 1106–1114.

OLVER, F.W.J. (1974), *Introduction to Asymptotics and Special Functions*, Academic Press, New York.

PARASNIS, D.S. (1979), *Principles of Applied Geophysics*, Chapman and Hall, London, 3 ed.

PARKER, R.L. (1977), 'Understanding inverse theory', *Annual Review of Earth and Planetary Sciences*, Vol. 5, Pp. 35–64.

PETERSON, B. and STRÖM, S. (1975), 'Matrix formulation of acoustic scattering from multilayered scatterers', *Journal of the Acoustical Society of America*, Vol. 57, No. 1, January, Pp. 2–13.

PETERSON, J.E., PAULSSON, B.N.P. and MCEVILLY, T.V. (1988), 'Applications of algebraic reconstruction techniques to crosshole seismic data', *Geophysics*, Vol. 50, No. 10, October, Pp. 1566–1580.

POGGIO, T., TORRE, V. and KOCH, C. (1985), 'Computational vision and regularisation theory', *Nature*, Vol. 317, No. 26, September, Pp. 314–319.

POTMESIL, M. (1987), 'Generating octree models of 3D objects from their silhouettes in a sequence of images', *Computer Vision, Graphics and Image Processing*, Vol. 40, Pp. 1–29.

PRATT, R.G. and WORTHINGTON, M.H. (1988), 'The application of diffraction tomography to cross-hole seismic data', *Geophysics*, Vol. 53, No. 10, October, Pp. 1284–1294.

RADON, J. (1917), 'On the determination of functions from their integral values along certain manifolds', *Berichte der Sachsischen Akademie der Wissenschatt*, Vol. 69, April, Pp. 262–277. Translation from German by P. C. Parks in *IEEE Transactions on Medical Imaging*, MI-5(4):170–176, December 1986.

RALL, L.B. (1969), *Computational Solution of Nonlinear Operator Equations*, Wiley, New York.

RICHARDS, W.A., KOENDERINK, J.J. and HOFFMAN, D.D. (1987), 'Inferring three-dimensional shapes from two-dimensional silhouettes', *Journal of the Optical Society of America A*, Vol. 4, No. 7, July, Pp. 1168–1175.

ROBB, R.A. (Ed.) (1985), *CRC Three-Dimensional Biomedical Imaging*, Vol. 1, CRC Press, Boca Raton, Florida.

ROGER, A. (1981), 'Newton-Kantorovitch algorithm applied to an electromagnetic inverse problem', *IEEE Transactions on Antennas and Propagation*, Vol. AP-29, No. 2, March, Pp. 232–238.

ROSENFELD, A. (1984), 'Image analysis: problems, progress and prospects', *Pattern Recognition*, Vol. 17, No. 1, Pp. 3–12.

ROSENFELD, A. (1988), 'Computer vision: Basic principles', *Proceedings IEEE*, Vol. 76, No. 8, August, Pp. 863–868.

SABATIER, P.C. (Ed.) (1978), *Applied Inverse Problems*, Lecture Notes in Physics volume 85, Springer-Verlag, Berlin. Lectures presented at the RCP 264 "Etude Interdisciplinaire des Problèmes Inverses" sponsored by the Centre National de la Recherche Scientifique.

SABATIER, P.C. (1983), 'Theoretical considerations for inverse scattering', *Radio Science*, Vol. 18, No. 1, January-February, Pp. 1–18.

SANTOSA, F., PAO, Y.H., SYMES, W.W. and HOLLAND, C. (Eds.) (1984), *Proceedings of International Conference on Inverse Problems of Acoustic and Elastic Waves, Cornell University*, Philadelphia.

SANZ, J.L.C. (1988), 'Computing image texture features in parallel computers', *Proceedings IEEE*, Vol. 76, No. 3, March, Pp. 292–294.

SANZ, J.L.C. and HINKLE, E.B. (1988), 'Note on contrast measures and polynomial classifiers', *Proceedings IEEE*, Vol. 76, No. 3, March, Pp. 294–296.

SCHOENBERGER, M. (Ed.) (1984), 'Special issue on seismic signal processing', *Proceedings IEEE*, Vol. 72, No. 10, October, Pp. 1235–1412.

SCHOMBERG, H. (1978), 'An improved approach to reconstructive ultrasound tomography', *Journal of Physics D: Applied Physics*, Vol. 11, Pp. L181–L185.

SCHULTZ, K.I. and JAGGARD, D.L. (1987), 'Microwave projection imaging for refractive objects: a new method', *Journal of the Optical Society of America A*, Vol. 4, No. 4, Pp. 1773 – 1782.

SENGBUSH, R.L. (1983), *Seismic Exploration Methods*, International Human Resources Development Corporation, Boston.

SHANKAR, P.M. (1986), 'Speckle reduction in ultrasound B-scans using weighted averaging in spatial compounding', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-33, No. 6, November, Pp. 754–764.

SLANEY, M., KAK, A.C. and LARSEN, L.E. (1984), 'Limitations of imaging with first-order diffraction tomography', *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-32, No. 8, August, Pp. 860–874.

SLEEMAN, B.D. (1982), 'The inverse problem of acoustic scattering', *IMA Journal of Applied Mathematics*, Vol. 29, Pp. 113–142.

SMITH, A.M., GOLDBERG, M. and LIM, E.S.K. (1980), 'Numerical ray tracing in media involving continuous and discrete boundaries', *Ultrasonic Imaging*, Vol. 2, No. 4, October, Pp. 291–301.

SOUMEKH, M. (1986), 'An improvement to the Rytov approximation in diffraction tomography', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-33, No. 4, July, Pp. 394–401.

SOUMEKH, M. and KAVEH, M. (1986), 'A theoretical study of model approximation errors in diffraction tomography', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-33, No. 1, January, Pp. 10–20.

SPENCER, A.J.M. (1980), *Continuum Mechanics*, Longman, London.

STRATMEYER, M.E. and LIZZI, F.L. (Ed.) (1986), 'Special issue on the biological effects of ultasound', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-33, No. 2, March.

STRÖM, S. and ZHENG, W. (1987), 'Basic features of the null field method for dielectric scatterers', *Radio Science*, Vol. 22, December, Pp. 1273–1281.

STRÖM, S. and ZHENG, W. (1988), 'The null field approach to electromagnetic scattering from composite objects', *IEEE Transactions on Antennas and Propagation*, Vol. AP-36, No. 3, March, Pp. 376–382.

TAN, D.G.H. (1985), *Implementable Multi-Dimensional Inverse scattering theory*, PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand.

TAN, D.G.H., QU, J.X. and BATES, R.H.T. (1986), 'Image reconstruction from projections VIII: effects of finite resolution and sampling of individual projections', *Optik*, Vol. 73, No. 1, September, Pp. 25–29.

TAN, D.G.H., MURCH, R.D. and BATES, R.H.T. (1989), 'Inverse scattering for penetrable obstacles', In *Proceedings of the 1989 URSI International Symposium on Electromagnetic Theory*, Royal Institute of Technology, Stockholm, Sweden, August 14-17, Pp. 160–162.

TAYLOR, K.J.W. (1979), 'Basic principles of diagnostic ultrasound', In K. PRESTON, K. J. W. TAYLOR, S.J. and AYERS, W.R. (Eds.), *Medical Imaging Techniques - A comparison*, Plenum Press, New York and London, Pp. 231–260.

TREVES, F. (1980), *Introduction to Pseudodifferential and Fourier Integral operators*, Vol. 1 - 2, Plenum, New York.

TYALOR, J.R. (1972), *Scattering theory*, Wiley, New York.

UFIMTSEV, Y. (1971), *Method of Edge Waves in the Physical Theory of Diffraction*, Sovyetskoye radio, Moscow. Translated by U.S. Air Force Foreign Technology Division, Wright-Patterson AFB, Ohio, USA.

VAN DYCK, D. (1987), Private commumication.

VAN HEEL, M. (1987), 'Angular reconstitution: A posteriori assignment of projection directions for 3D reconstuction', *Ultramicroscopy*, Vol. 21, Pp. 111–124.

VARADAN, V.K., VARADAN, V.V. and PAO, Y.H. (1978), 'Multiple scattering of elastic waves by cylinders of arbitrary cross section', *Journal of the Acoustical Society of America*, Vol. 63, No. 5, May, Pp. 1310–1319.

VARATHARAJULU, V. and PAO, Y.H. (1976), 'Scattering matrix for elastic waves. I. Theory', *Journal of the Acoustical Society of America*, Vol. 60, No. 3, September, Pp. 556–566.

WAIT, J.R. (1962), *Electromagnetic Waves in Stratified Media*, Vol. 3 of International series of monographs on Electromagnetic waves, Pergamon Press, New York.

WALL, D.J.N. (1980), 'Methods of overcoming numerical instabilities associated with the T-matrix method', In VARADAN, V.K. and VARADAN, V.V. (Eds.), *Acoustic, Electromagnetic and Elastic Wave Scattering — Focus on the T-matrix approach*, Pergamon Press, New York, Pp. 269–286.

WALL, D.J.N., YEO, T.S. and BATES, R.H.T. (1985), 'Inverse scattering and the null field method', In DEVANEY, A.J. and BATES, R.H.T. (Eds.), *Inverse Optics II*, August, Pp. 25–29. Proceedings SPIE volume 558 (the Society of Photo-Optical Instrumentation Engineers).

WALTHER, A. (1988), 'Numerical techniques in Eikonal function theory', *Journal of the Optical Society of America A*, Vol. 5, No. 4, April, Pp. 511–514.

WANG, Y.F., MAGEE, M.J. and AGGARWAL, J.K. (1984), 'Matching three-dimensional objects using silhouettes', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 4, July, Pp. 513–518.

WATERMAN, P.C. (1969), 'New formulation of acoustic scattering', *Journal of the Acoustical Society of America*, Vol. 45, No. 6, June, Pp. 1417–1429.

WATSON, G.N. (1966), *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, Cambridge, second ed.

WEISS, I. (1988), '3D shape representation by contours', *Computer Vision, Graphics and Image Processing*, Vol. 41, No. 1, January, Pp. 80–100.

WELLS, P.N.T. (1977), *Biomedical Ultrasonics*, Academic Press, London.

WESTCOTT, B.S. (1983), *Shaped reflector antenna design*, Research studies press, England.

WILCOX, C.H. (1984), *Sound Propagation in Stratified Fluids*, Springer-Verlag, New York. Applied Mathematical Sciences volume 50.

WITKIN, A.P. (1981), 'Recovering surface shape and orientation from texture', *Artificial Intelligence*, Vol. 17, August, Pp. 17–45.

WOUK, A. (1979), *A Course of applied functional analysis*, Wiley-interscience, New York.

WU, R. and TOKSÖZ, M.N. (1987), 'Diffraction tomography and multisource holography applied to seismic imaging', *Geophysics*, Vol. 52, No. 1, january, Pp. 11–25.

YEO, T.S. (1985), *Diffraction by a Penetrable Wedge*, PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand.